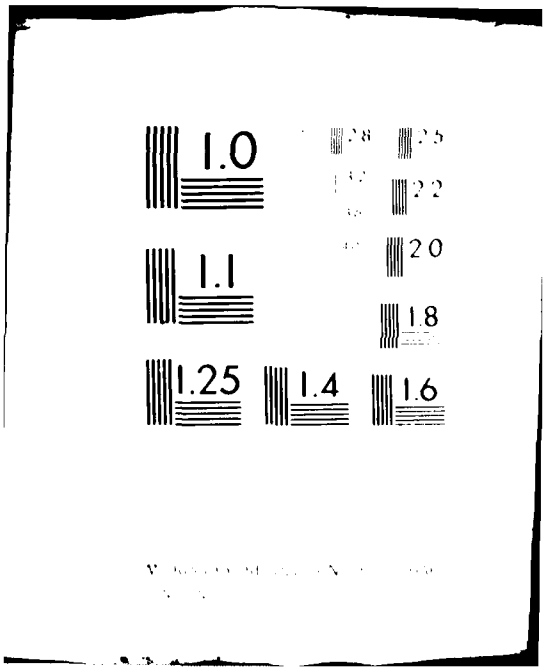
AD-A089 938 IBM THOMAS J WATSON RESEARCH CENTER YORKTOWN HEIGHTS NY F/G 20/12
THEORETICAL STUDIES OF DEFECTS IN TETRAHEDRAL SEMICONDUCTORS,(U)
AUG 80 S T PANTELIDES, J A VANVECHTEN F49620-79-C-0077
UNCLASSIFIED AFOSR-TR-80-0758 NL

1 OF 2


1.0
1.1
1.25
1.4
1.6
2.2
2.0
1.8

AFOSR-TR-80-0758

THEORETICAL STUDIES OF DEFECTS IN TETRAHEDRAL SEMICONDUCTORS

Sokrates T. Pantelides and James A. VanVechten

IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, New York 10598

August, 1980

Final Technical Report for the period June 15, 1979 - June 14, 1980
Contract No. F49620-79-C-0077

Approved for public release, distribution unlimited

Prepared for:

Air Force Office of Scientific Research (AFSC)
Bldg. 410, Bolling AFB, D.C. 20332

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-80-0758	2. GOVT ACCESSION NO. AD-A089938	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Theoretical Studies of Defects in Tetrahedral Semiconductors		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) S. T. Pantelides and J. A. VanVechten		8. CONTRACT OR GRANT NUMBER(s) F49620-79-C-0077
9. PERFORMING ORGANIZATION NAME AND ADDRESS IBM T. J. Watson Research Center P. O. Box 218 Yorktown Heights, NY 10598		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2301/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bldg. 410 N/P Bolling AFB, DC 20332		12. REPORT DATE August 1980
		13. NUMBER OF PAGES 128
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Defects, Semiconductors, Dislocations, Impurities

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is the final report on work that has been carried out on the electronic and thermochemical properties of impurities, point defects and dislocations in semiconductors.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT AFOSR CONTRACT

No. F49620-79-C-0007

A. DEFECT THERMOCHEMISTRY

Contract work in this area has focussed on two areas - the mechanism of pulsed laser annealing and the proper analysis of deep level transient spectroscopy (DLTS) spectra.

The pulsed laser annealing mechanism has been a controversial subject. Although early workers, particularly in the U.S.S.R., felt it must be a nonthermal mechanism of some sort, the opinion, originated by E. Rimini of Italy, that it was purely thermal involving ordinary melting and very rapid recrystallization became widely accented in the U.S.A. and most other Western countries. Together with collaborators in the Netherlands, we showed that most experimental data regarding the pulsed annealing, particularly acoustic, optical reflectivity, impurity redistribution, and surface strain data could not reasonably be attributed to a strictly thermal process with severe temperature gradients and transients (Appendix A, B, C). We developed the alternative theory that annealing was occuring in the presence of a dense electron-hole plasma produced by the laser and persisting as long 800 ns. (Appendix B, C, D) with lattice temperature no higher than 300°C. We treated the kinetics of defect motion and removal under such conditions, the diffusion of the plasma and reasons that it would persist so much longer than had previously been believed (Appendix C, D, E, F, G, H). striking confirmation of our plasma-induced annealing theory has now been obtained by H. W. Lo and A. Compaan, Phys. Rev. Lett 44, 1604 (1980) and to be presented at the September Conferences in Japan. With time-resolved Raman measurement of surface lattice temperature, they confirmed that it was less than 300°C during pulsed laser annealing condition, at least up to 40ns after the end of the annealing pulse, while the carrier temperature remained over 2000°C.

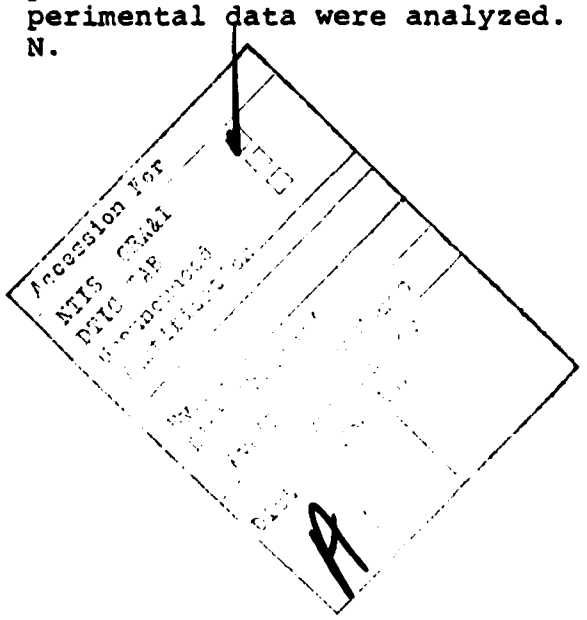
In the DLTS analysis, we sought to understand why the reported defect concentrations were less than predicted by theory, observed by transmission electron microscopy and thought to be required in order to explain carrier lifetimes and the degradation of electro-optic devices. The conventional DLTS analysis assumes that the experiment succeeds in trapping only electrons or holes at the defect sites and then measures only the detrapping of those carriers singley when the material is driven into depletion. We showed (Appendix I) that, in addition to this contribution to the DLTS signal, there are also contributions from recombination, reorientation, and internal reconfiguration of multiple carriers at defects that are composed of two or more point defects. However, these multiple contributions are much weaker for each event

than simple detrapping. Thus, the number of such defects contributing to a given experimental spectrum would be much larger than the conventional analysis implies. We also show that in several cases it is likely the present experimental design would not succeeed in screening out such multiple carrier effects. We suggest experimental tests of this hypothesis; some data in support of it has already been obtained.

B. ELECTRONIC STRUCTURE

The major accomplishment in the area of electronic structure is the self-consistent calculation of the reconstructed vacancy in Si. After completing our work on the unrelaxed vacancy, described in detail in Appendices J and K, we extended the method and calculated the electronic structure which has undergone a Jahn-Teller distortion as determined by experiments. The results of the calculation are described in Appendix L. In independent work, we further extended the method to treat chemical impurities. As part of this contract, comparisons were drawn between the results of the vacancy and the impurities and understanding of both was furthered. (Appendix M.)

Another issue that was addressed is the nature of final stages in the process of ionizing deep defects and impurities. The importance of localized final states was stressed and available experimental data were analyzed. The details are given in Appendix N.



Appendix A

Volume 74A, number 6 PHYSICS LETTERS 10 December 1979

REASONS TO BELIEVE PULSED LASER ANNEALING OF Si DOES NOT INVOLVE SIMPLE THERMAL MELTING

J.A. Van VECHTEN [1] and R. TSU
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

and

F.W. SARIS and D. HOONHOUT
FOM Instituut voor Atoom- en Molecuulfysica, Amsterdam, The Netherlands

Received 3 September 1979

Many recent publications dealing with Si pulsed laser annealing have assumed that the transformation is strictly thermal melting–recrystallization. We recount observations indicating the material was not subjected to thermal melting.

In many of the recent publications dealing with the annealing of amorphous Si by pulsed (30 ps to 150 ns) lasers operating in the 1.06 to 0.26 μm wavelength region, it has been claimed that the annealing transformation occurs by a strictly thermal melting and recrystallization mechanism. It has been assumed: (a) that the energy of the incident photons is converted into heat in the lattice in a layer of depth given by the photon absorption length, and in a time small compared to the duration of the pulse; (b) that, to obtain a good anneal, it is necessary to melt through the amorphous layer to the single crystal Si substrate; and (c) that the recrystallization process, which occurs as the material cools, is liquid phase epitaxial growth [1–4].

However, in a recent review of the subject [5], Khaibullin, who is credited with the discovery of pulsed laser annealing, concluded "The mechanism of laser annealing is not yet finally established. But even now one can state that in the case of the nanosecond regimes one cannot reduce it to the ordinary thermal effect. Different factors such as photoionization, impact wave, powerful light fields etc. play a significant role". Two particularly convincing reasons were given for dismissing the melting hypothesis. First, crystalline order at the surface was observed by diffraction techniques to occur within 20 to 30 ns of the beginning of a 15 to 20 ns annealing pulse. If melting had occurred, then by all estimates the molten phase would have required several times this interval to recrystallize. Second, as the molten phase has a reflectivity at normal incidence that varies between 72 and 74% through the visible wavelengths whereas amorphous and crystalline Si have significantly lower reflectivities at all temperatures, an abrupt increase in the amount of radiation reflected back should have been observed as soon as the sample surface had melted about one skin depth deep (about 8 nm), but this was repeatedly observed not to occur in cases where full annealing was observed. The second observation has also been made in experiments in Japan [6] and in our laboratory [7].

The simple thermal melting model is also inconsistent with at least five other observations.

(1) Let us continue on the subject of the reflectivity of the sample surface during the annealing event. In several laboratories a significant increase in the reflectivity of the surface has been observed [6–10] at wave lengths ranging from 1.06 μm to 533 nm, but not at 266 nm, for normal incidence and for both 57.5° and 45° off normal incidence. Let us consider how well the observations of this reflectivity increase

[1] Supported in part by the Air Force Office of Scientific Research under Contract No. F49620-79-C-0077.

fit the hypothesis that it is the result of the presence of a layer of ordinary molten Si at the surface of the sample. The complex index of refraction of molten Si has been measured by Shvarev et al. [11] [*1] at 1.0, 0.7 and 0.4 μm from 57.5° off normal incidence. Auston et al. [8] reported the time resolved reflectivity at $\lambda = 0.63$ μm, 57.5° from the surface normal for both parallel and perpendicular polarization, $R_\parallel$ and $R_\perp$, during laser annealing at 1.06 μm with 50 ns pulses. The observed values for the $R_\parallel$ and $R_\perp$ during the high reflectivity phase of the annealing process were $R_\parallel$ (0.63 μm) = 47% and $R_\perp$ (0.63 μm) = 82%, respectively. The values for molten Si are $R_\parallel$ (0.7 μm) = 57%, $R_\parallel$ (0.4 μm) = 56% and $R_\perp$ (0.7 μm) = 85%, $R_\perp$ (0.4 μm) = 84%. One might imagine several effects, including the presence of a vapor over the surface or ripples on the surface or a temperature dependence of the reflectivity, which might cause the reflectance measured during pulsed laser annealing to differ from that normally measured from molten Si. However, most of these would affect both $R_\perp$ and $R_\parallel$ whereas we see that the discrepancy in the former is only 2% while that of the latter is 20%. If the temperature of the surface were affecting the measurement, then at the end, just before recrystallization, the reflectivity should have risen to the normal value for molten Si at the melting point.

When one considers the time dependence of the reflectivity, two qualitative discrepancies between the data and the thermal melting hypothesis are evident. At the beginning of the high reflectivity phase, both Auston et al. [8] and Hodgson et al. [7] observed that the reflectivity increase occurred near the peaks of their pulses at the lowest power for which it was observed. (The rise was quite abrupt and moved out of the leading edge of the pulse as the net pulse energy was increased.) If the reflectivity increase were due to the presence of a molten layer, or due to any process that simply involved heating the surface layer to a high temperature, then this increase would have occurred well down the trailing edge of the pulse at the lowest total energies. This is obvious if one considers the case of a square wave pulse for which the maximal temperature must always occur at the end of the pulse. One can easily prove that the maximal temperature can never occur at the peak or leading edge for any shaped pulse. The exact time of the maximal surface temperature depends on pulse duration, thermal diffusivity (which is generally T dependent), pulse shape, and extinction coefficient of the laser light. In general, the shorter the pulse duration, the squarer the pulse shape, or the lower the thermal diffusivity, the farther out of the trailing edge the maximal temperature occurs. This is because less of the integrated absorbed laser power diffuses away. Using the diffusivity of heat in Si measured under furnace conditions, one finds [*2] that for the conditions of Auston et al.'s experiment [8,9] the onset of melting for a minimum energy pulse that melts one skin depth deep would occur 30 to 40 ns after the peak of their 25 ns HWHM gaussian pulse rather than within experimental error (less than 5 ns) of the peak. One may force the onset of the reflectivity rise to agree with experiment by adjusting the assumed value of the diffusivity of whatever it is that causes the reflectivity to rise at a critical density. In order to bring the calculated position of the reflectivity rise into agreement with experiment, one must assume the diffusivity is at least two orders of magnitude larger, e.g. 18 cm^2/s, the ambipolar diffusivity of a plasma in Si at room temperature [13], instead of 0.117 cm^2/s, the thermal diffusivity at the melting point [14,15].

When the reflectivity decreases back to values appropriate to normal single crystal Si, this should, according to the melting hypothesis, occur in a time less than 5 ns. (The skin depth is less than 10 nm and estimates of the recrystallization rate range [1–3] from 2.5 m/s to 4.5 m/s and even higher.) However, the observation, for example fig. 1 of the reports by Auston et al. [8,9] is that the duration of this transition is about 100 ns. Although Auston et al. [8] were sampling a spot centered in their annealed region and containing less than 1% its area, one might suggest that inhomogeneity in the laser beam caused the material to melt much deeper in some places than in others and that this might explain the slow decay of the reflectance. Such an effect would have produced a

[*1] Cf., e.g., Stratton [12] for the relation between complex index of refraction and reflectivity of a metallic surface off normal incidence.

[*2] We are grateful to G.J. Lasher for providing the computer simulation code used to establish this point. A similar conclusion for somewhat different laser annealing conditions is drawn from fig. 3 of Baeri et al. [2].

rough surface with uneven dopant concentration both laterally and vertically. Good quality material does not show this [16].

(2) Consider now the mechanical forces that must act on the surface layer if it were to melt. Suppose that an amorphous layer 100 nm thick were to be heated to the melting point, 1400°C. in 10 ns. The outward acceleration due to thermal expansion [17] ($\alpha = 4 \times 10^{-6}$/°C) would be 5.6×10^8 cm/s^2 = 6×10^5 G. If this layer were to melt in the next 10 ns, then, due to the 15% contraction upon melting [‡3], there must be a snap-back force and an acceleration back towards the substrate if the molten layer is not to fly off. If this contraction were three-dimensional (as seems to us most likely) then that acceleration would be 5×10^6 G, but this would require that the surface breakup into an array of small molten drops, which would seem inconsistent with the surface morphology flat to within 5 nm found by various authors after annealing. If, in order to account for this morphology, we assume the contraction to be one dimensional, then the snap-back acceleration must be 1.5×10^7 G. The only source of such a snap-back force that we can imagine is the surface tension of the molten phase, but the surface tension acts only in the plane of the surface. Surface tension can only hold the liquid in contact with the substrate in places where the liquid is "balled up" so that the surface is substantially normal to the substrate and extends down to it. But this would not produce the very flat annealed region that is observed.

(3) Von Gutfeld and Tsu [21] have measured the thermoelastic waves produced by the incident laser beam on the free surface of the Si during the pulsed laser annealing event as a function of incident energy. These waves were detected by a 20 MHz ± 25% contact transducer attached to the opposite face of the Si wafer. The elastic signal is observed to rise smoothly with increasing incident energy density. There is no discontinuity at the threshold for single shot annealing nor in the range in which good quality material is obtained. There is a sharp discontinuity at higher powers for which surface damage and evidence of Si evaporation are observed. This discontinuity is not unlike that observed by McCelland and Kniseley [22] at the point where Ge was melted with a chopped CW laser. These observations of von Gutfeld and Tsu are consistent with crystallization without thermal melting and the concomitant volume contraction. Further experiments are in progress to calibrate and elaborate this thermoelastic technique.

(4) Baeri et al. [23] have irradiated Cu-implanted Si crystals with Q-switched ruby laser single pulses. After irradiation at an energy density in excess of 1 J/cm^2, Cu was found to accumulate at the surface. Baeri et al. argue that accumulation of Cu atoms is a consequence of surface layer melting and resolidification, the redistribution being related to the low value (4×10^{-4}) of the segregation coefficient of Cu. Similar arguments were put forward by Cullis et al. [24], who observed redistribution of Ga and by White et al. [25], for Cu and Fe. However, in the last case Fe was found to segregate to the surface less than Cu although the redistribution coefficient of Fe is smaller.

Hoonhout and Saris [26] have made a systematic investigation of this effect for eight common dopants in Si with redistribution coefficients ranging from 0.33 to 10^{-8}. Their results show that, under implantation and pulsed laser annealing conditions similar to those of Baeri et al. [23], there is no correlation between the observation of dopant redistribution and the normal thermal redistribution coefficient. Bi, which has a redistribution coefficient $k_0 = 7 \times 10^{-4}$ similar to that of Cu, does not segregate as much as Cu. Se and Te ($k_0 = 10^{-8}$ and 8×10^{-6}, respectively) are observed to segregate even less. Assuming that impurities are frozen in because the liquid–solid interface moves to the surface faster than the impurities, one would still predict those impurities which have the lowest redistribution coefficient to segregate most. Yet this is not observed, whereas liquid-phase diffusivities for these impurities do not vary by more than one order of magnitude. Hence, surface segregation cannot be used as evidence for the thermal melting model.

(5) The fact that the "Heating 5" computer simulation code [1] of Oak Ridge National Laboratory gives a reasonable fit to the redistribution data for a number of cases actually militates against the melting hypothesis because this code neglects five large effects which would greatly increase the calculated redistribution if

[‡3] Van Vechten [18, especially p. 1491] deduced this value from the Clapeyron equation, the entropy of fusion of Hultgren et al. [19] and the phase diagram of Bundy [20]. A value of 10%, based on an old volumetric measurement, is sometimes quoted, but we regard this value as less reliable due to experimental difficulties with molten Si.

included. (Actually, for the specific case presented in ref. [1], Heating 5 has already overestimated the redistribution. It was calculated that the doped layer would remain molten for 250 ns, whereas the distribution data was fit by assuming it was molten for only 180 ns.) The five effects neglected by Heating 5 are as follows: (1) As shown by van Gurp et al. [27], the molten layer should become turbulent in times short compared with the calculated persistence of the molten phase. The Heating 5 calculation assumed there was no turbulence. Turbulence will greatly increase the redistribution of elements within the molten layer and also produce lateral inhomogeneities. (2) The impurity diffusivity, $D^{\mathrm{L}}(\mathrm{I})$, is assumed constant and equal to that measured at the melting point of pure Si, T^{F}, although the calculation shows the temperature of the liquid going 1400 K above T^{F}. It is known that liquid state diffusivities are temperature dependent [28] and usually well described by an activation energy, $\Delta H_{\mathrm{m}}(\mathrm{I})$, of order 0.5 eV. Therefore, an order of magnitude increase in the effective $D^{\mathrm{L}}(\mathrm{I})$ should have been assumed. (3) In fact the initial material is not single crystal but is amorphous and thus possesses an excess heat of recrystallization $\Delta H^{\mathrm{R}}(\mathrm{Si})$ which has been estimated by Turnbull [29] to be

$$1/3\,\Delta H^{\mathrm{F}}(\mathrm{Si}) = \Delta H^{\mathrm{R}}(\mathrm{Si}) > 0.5\,(T^{\mathrm{F}} - 300\ \mathrm{K})\,C_p\,,$$

where ΔH^{F} and T^{F} are the heat and temperature of fusion and C_p is the specific heat. This effect lowers the energy required to melt the damaged layer by an amount about half as large as that required to raise the temperature of an equivalent mass of pure single crystal Si to its melting point. The material should melt sooner and deeper than calculated. (4) The suppression of T^{F} due to the impurity I was neglected. The magnitude of the suppression of the equilibrium freezing point $T^{\mathrm{F}}(\mathrm{Si}, \mathrm{I}, X)$ for an impurity concentration, X, can be found in the literature [30] up to the normal (equilibrium) saturation value X_{s}. Most values, e.g. for I = As, are

$$T^{\mathrm{F}}(\mathrm{Si}) - T^{\mathrm{F}}(\mathrm{Si}, \mathrm{I}, X_{\mathrm{s}}) \approx 200\ \mathrm{K}\,.$$

However, much of the empirical data being fit are for concentrations grossly in excess of normal saturation, $X \gg X_{\mathrm{s}}$. For example in the case of I = B, $X = 10^4 X_{\mathrm{s}}$ is reported. In this circumstance $T^{\mathrm{F}}(\mathrm{Si}) - T^{\mathrm{F}}(\mathrm{Si}, \mathrm{I}, X_{\mathrm{s}})$ is certainly very large and would require a detailed knowledge of the kinetics of the particular reaction to determine. The material would melt sooner and stay molten longer than Heating 5 would calculate. (5) The degree of supercooling required to make the material recrystallize at the very large rate (about 200 cm/s) assumed is very large. In view of the considerations just mentioned, no rigorous estimate of this effect can be made. However, one could note that float zone or Czochralski Si is normally grown at about 3 mm/min $= 5 \times 10^{-3}$ cm/s with about 0.5 K of supercooling [31].

In conclusion, although we do not doubt that it is possible to melt Si with a laser pulse, we claim that those annealed samples having the remarkably high crystallographic and morphological quality are probably obtained by pulsed laser annealing via a nonthermal process in which thermal gradients are mild at least at the time that rapid atomic motion ceases. We believe that a realistic description of pulsed laser annealing must take account of the dense plasma produced by the absorption of the light. This will be discussed in the accompanying paper [32].

References

[1] J.C. Wang, R.F. Wood and P.P. Pronko, Appl. Phys. Lett. 33 (1978) 455.

[2] P. Baeri, S.U. Campisano, G. Foti and E. Rimini, J. Appl. Phys. 50 (1979) 788.

[3] C.M. Surko, A.L. Simons, D.H. Auston, J.A. Golovchenko and R.E. Slusher, Appl. Phys. Lett. 34 (1979) 635.

[4] J.C. Schultz and R.J. Collins, Appl. Phys. Lett. 34 (1979) 363.

[5] I.B. Khaibullin, E.I. Shtyrkov, M.M. Zaripov, R.M. Bayazitov and M.F. Galjautdinov, Radiat. Eff. 36 (1978) 225;
see also G.A. Kachurin, V.A. Bogatyriov, S.I. Romanov and L.S. Smirnov, in: Ion implantation in semiconductors (1976), eds. F. Chernow, J.A. Borders and D.K. Brice (Plenum, New York, 1977) p. 445.

[6] K. Murakami, M. Kawabe, K. Gamo, S. Namba and Y. Aoyagi, Phys. Lett. 70A (1979) 332.

[7] R.T. Hodgson, R. Tsu, J.A. Van Vechten and E.J. Yoffa, Bull. Am. Phys. Soc. 24 (1979) 315.

[8] D.H. Auston, C.M. Surko, T.N.C. Venkatesan, R.E. Slusher and J.A. Golovchenko, Appl. Phys. Lett. 33 (1978) 437.

[9] D.H. Auston, J.A. Golovchenko, A.L. Simons, C.M. Surko and T.N.C. Venkatesan, Appl. Phys. Lett. 34 (1979) 777.

[10] Y.S. Liu and K.L. Wang, Appl. Phys. Lett. 34 (1979) 363.

[11] K.M. Shvarev, B.A. Baum and P.V. Geld, Sov. Phys. Solid State 16 (1975) 2111.

[12] J.A. Stratton, Electromagnetic theory (McGraw-Hill, New York, 1941) p. 505.
[13] H. Wolf, Semiconductors (Wiley, New York, 1971) p. 290.
[14] H.R. Shanks, P.D. Maycock, P.H. Sidles and C.G. Danielson, Phys. Rev. 130 (1963) 1743.
[15] C.J. Glassbrenner and G.A. Slack, Phys. Rev. 134 (1964) A1058.
[16] C.W. White, J. Naranayan and R.T. Young, Science 204 (1979) 461.
[17] H. Ibach, Phys. Stat. Sol. 31 (1969) 625.
[18] J.A. Van Vechten, Phys. Rev. B7 (1973) 1479.
[19] R. Hultgren, R.L. Orr, P.D. Anderson and K.K. Kelley, Selected values of thermodynamic properties of metals and alloys (Wiley, New York, 1963).
[20] F.P. Bundy, J. Chem. Phys. 41 (1964) 3809.
[21] R. Tsu, J.E. Baglin, T. Tan and R.J. von Gutfeld, Electrochem. Soc. Extended Abstr. 79-2 (1979).
[22] J.F. McCelland and R.N. Kniseley, Appl. Phys. Lett. 35 (1979) 121.
[23] P. Baeri, S.U. Campisano, G. Foti and E. Rimini, Phys. Rev. Lett. 41 (1978) 1246.
[24] A.G. Cullis, J.H. Poate and G.K. Celler, in: Laser–solid interactions and laser processing, eds. S.D. Ferris, H.J. Leamy and J.H. Poate, AIP Conf. Proc. No. 50 (1979) p. 311.
[25] C.W. White, J. Narayan, B.R. Appleton and S.R. Wilson, J. Appl. Phys. 50 (1978) 2967.
[26] D. Hoonhout and F.W. Saris, Phys. Lett. 74A (1979) 253.
[27] G.J. Van Gurp, G.E.J. Eggermont, Y. Tamminga, W.T. Stacy and J.R.M. Gijsbers, Appl. Phys. Lett. 35 (1979) 273.
[28] D. Turnbull, private communication; see also H.S. Chen and D. Turnbull, J. Appl. Phys. 40 (1969) 4214.
[29] K. Kakeko, M. Ayabe, M. Dosen, K. Morizane, S. Usui and N. Watanabe, J. Electrochem. Soc. 121 (1974) 556.
[30] W.R. Runyan, Silicon semiconductor technology (McGraw-Hill, New York, 1965) p. 266.
[31] A.J.R. de Kock, P.J. Roksnoer and P.G.T. Boonen, Semiconductor silicon (1973), eds. H.R. Huff and R.R. Burgess (The Electrochem. Soc., Princeton, 1973) p. 83.
[32] J.A. Van Vechten, R. Tsu and F.W. Saris, following paper.

JOURNAL DE PHYSIQUE

Colloque C4, *supplément au n°* 5, *Tome* 41, *mai* 1980, *page* C4-15

Appendix B

IMPORTANCE OF THE PLASMA TO PULSED LASER ANNEALING (*)

J. A. Van Vechten

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, U.S.A.

Abstract.- In the presence of a dense (> $10^{21}/cm^3$) electron-hole plasma which may be produced by the intense irradiation during pulsed laser annealing, the covalent bonding of tetrahedral semiconductors like Si will be severely weakened because a significant fraction of the bond charge has been excited across the gap into anti-bonding or plane wave states. The crystal structure may even become unstable and undergo a phase transition if the bond charge is reduced beyond a critical point where the transverse acoustic modes go to zero frequency. This transition has previously been estimated from the temperature dependence of the gap to occur at a density of $8\times10^{21}/cm^3$ for a lattice temperature of 0 K and at lower densities for higher lattice temperatures. In this state the material would not be able to support a shear and so should be called a liquid, but it would be very distinct from molten Si. The energy from the laser would be concentrated in the electronic excitation and not in kinetic energy of the atoms, as in the thermal case. Recrystallization could occur without the destructive effects of severe thermal gradients when the material passes back through this phase transition to the covalently bonded phase. However, recrystallization of ion-implanted material should also be possible without passing through this phase transition if the atoms are reordered by the optically induced gliding of dislocations out of the damaged zone, as has been observed in crystalline semiconductors. Point defects should be removed from amorphous or damaged material at a rate much greater than normal furnace annealing due to five separate effects. First, they may be eliminated by recrystallization following the phase transition. Second, they may experience recombination enhanced diffusion within the covalent phase. Third, the coulombic trapping by charged impurities, which normally reduces the net rate of vacancy (or interstitial) migration drastically, will be suppressed by the flux of free carriers in the dense plasma. Fourth, point defects may be swept out of the material as the large concentration of dislocations, which are ideal sinks, glide out. Fifth, the energy of formation of fast diffusing interstitial species can be greatly reduced by the dense plasma. It is claimed that these effects are required to account for the very high crystallographic, morphological and electrical quality of Si produced by pulsed laser annealing.

In his 1977 review of the subject of pulsed laser annealing I. B. Khaibullin concluded /1/, "The mechanism of laser annealing is not yet finally established. But even now one can state that in the case of the nanosecond regimes one cannot reduce it to the ordinary thermal effect. Different factors such as photoionization, impact wave, powerful light fields etc. play a significant role." Kachurin, et al., /2,3/ also showed evidence that pulsed laser annealing is a nonthermal process.

In addition to the convincing arguments that pulsed laser annealing could not be a simple thermal effect and could not be simple thermal melting followed by rapid crystallization and quench which are given in those papers, several more observations and arguments lead to the same conclusion. There is, of course, the fact that long standing theories for laser damage /4-7/ indicate that the material should be heavily damaged, not annealed, if all the energy of the laser pulse were converted to local heating of the material in times of order the duration of the pulse. Simple analysis /8/ of the mechanical forces that would be generated by thermal expansion and by contraction upon melting in times of order the duration of these laser pulses show that a molten layer could not remain in contact with the substrate unless it balled-up to an extent completely inconsistent with the very flat surface morphology that is obtained. Thermoelastic measurements /9/ made during pulsed laser annealing of Si show no evidence for the snap-back force that would have to occur if the surface melted and contracted until one increases the laser energy above the damage threshold. Indeed, it has been shown that thermal expansion forces in Si subjected to either 1 μs or 100 μs

* Supported in part by Air Force Office of Scientific Research under Contract No. F49620-79-C-0077.

pulses /10,11/ will fracture the material without raising its temperature to the melting point. In addition to the observation that annealing may occur without any significant increase in surface reflectivity /1, 12-14/ it has been noted /8/ that when a rise in surface reflectivity is observed /15-17/ it rises and falls in a manner expected for a dense electron-hole plasma and not that expected for molten Si. (Indeed, the traditional explanation for the abrupt rise in reflectivity of Si mirrors used for Q-switching lasers has been the formation of a sufficiently dense electron-hole plasma /18/). Moreover, there is no close agreement /8,13,14,16,19/ between the measured reflectivity during the high reflectivity period and the known reflectivity of molten Si /20/. There is close agreement between the reflectivity of molten Si and that measured when one increases the laser energy well above the threshold for damage to the wafer /14/. Furthermore, the redistribution of dopants has been shown /8,21,22/ to be inconsistent with the hypothesis of crystallization from normal molten Si.

It may also be noted that Matthews and Ashby observed /23/ nonthermal annealing of ion implanted amorphous Si by the electron beam of their electron microscope in 1972. The author believes the physical processes involved in electron beam annealing are essentially the same as those in laser annealing.

It is the author's opinion that it is the photoionization, the dense electron-hole plasma created by the intense laser pulse, which principally accounts for the remarkable properties of pulsed laser annealing and that the process is indeed nonthermal in the sense that the thermal transients that do occur are neither helpful to the annealing nor particularly severe. Let us call this the "Plasma Annealing", PA, theory and process.

By far the most difficult aspect of the PA theory of pulsed laser annealing is the question why the plasma remains dense for such a long time and transfers its energy to the lattice so slowly. However, it seems that the observations that the increased surface reflectivity, which has the properties of a dense plasma and not those of a molten layer, may persist /16,17/ as long as 800 ns demonstrate the point experimentally. Explanations for this long lifetime and slow transfer are given in detail elsewhere /24-27/. These explanations involve: a) the decoupling of the lattice from the plasmons of the excited carriers when $\hbar\omega_p$ becomes much larger than all phonon modes /25/ (when the excited carrier density exceeds $10^{19}/cm^3$); b) the screening of the deformation potential coupling between hot carriers and phonons /26/; c) an argument that under normal conditions the coupling of the lattice to the carriers via the plasmons is more important than that via the deformation potential, particularly for intervalley scattering with the emission of longitudinal phonons /24/; and d: a calculation showing that in Si the only optic phonons coupling to plasmons have (100) wave vectors so that plasmons with wave vectors in other directions will be long lived when $\hbar\omega_p$ is much larger than acoustic branch energies /27/.

However, let us note that even if there were no slowing of the energy transfer from hot carriers to the lattice, even if the phonon emission time were 1×10^{-13}s, one would not melt a typical 100 nm thick damaged layer back to the substrate at the observed threshold for annealing. The energy required to heat crystalline Si from room temperature to the melting point and to melt it is 7.3×10^3 J/cm^3. Due to the heat of recrystallization (which has not yet been determined experimentally for Si), the energy required to heat and melt amorphous Si is somewhat, perhaps a third, less. Therefore, the energy required to melt a 100 nm layer of ion implanted Si be about 0.05 J/cm^2 if there were no diffusion of energy to the substrate. As the normal incidence reflectivity rises from about 35 % at the beginning of the pulse to about 60 % during the observed high reflectivity period, the energy absorbed is about half that incident. Thus the incident energy density threshold required to melt a 100 nm layer with no diffusion would be about

0.10 J/cm^2 for an amorphous ion implanted surface. The observed incident energy threshold for single shot complete annealing of a 100 nm amorphous layer with either 30 ns or 30 ps pulses is about 0.5 J/cm^2 for 533 nm light /19,28/ which is absorbed in the same 100 nm or about 0.25 J/cm^2 for 266 nm light which is absorbed in the first 8 nm. The energy of a 533 nm photon exceeds the energy gap of Si by more than 1 eV; this energy would appear as kinetic energy of the carriers excited across the gap. Free carrier absorption would leave the carrier with more than 2.3 eV of kinetic energy. Auger recombination would add 1.1. eV to the kinetic energy of the third carrier in the process. Even if a phonon emission time of 1×10^{-13} s is assumed, these carriers would still diffuse more than 200 nm before giving up 1 eV to the lattice. This alone would double or triple the energy required to melt 100 nm with 533 or 266 nm light. The thermal diffusivity of Si under normal furnace conditions (in the dark) varies from 0.86 cm^2/s at room temperature to 0.117 cm^2/s at the melting point /29,30/. If we take D = 0.117 and τ = 30 ns, a typical laser pulse duration, and if we make a rough estimate of the extent of thermal diffusion, we find $(D\tau)^{1/2}$ = 600 nm or six times depth of the layer to be annealed. (For a more detailed account of these two effects see the recent study of E. J. Yoffa /31/) In fact the diffusion of the heat that has been delivered to the lattice will be substantially greater than this estimate because the laser pulse produces a carrier concentration which, by all estimates, is much larger than the thermal intrinsic value (2×10^{19}/cm^3 at the melting point and less at lower temperatures) and because 40 % of the thermal diffusivity at the melting point is due to the intrinsic free carriers /30/. If we make the conservative estimate /26/ that the carrier concentration during the laser pulse is 2×10^{20}/cm^3, then the diffusivity at the melting point would be 0.54 cm^2/s and $(D\tau)^{1/2}$ = 1.3 μm. Combining this consideration with that of the diffusion of the carriers even with a 1×10^{-13} s phonon emission time, we conclude that the incident energy would have to exceed 1 J/cm^2 to melt a 100 nm amorphous layer with either 533 or 266 nm light. Thus the damaged layer will not be melted back to the substrate at the observed incident energy density thresholds of 0.25 J/cm^2 for 266 nm or 0.5 J/cm^2 for 533 nm light. (Some published computer simulations /19,32-34/ of the heating of the material during pulsed laser annealing have concluded that the annealing threshold is the level at which melting does extend to the damage interface. These neglect the effects of carrier diffusion as phonons are being emitted and the enhancement of thermal diffusivity due to the excess carrier concentration produced by the intense laser light).

Now let us outline the PA hypothesis. Two central ideas are that the presence of a large number of excited electrons and holes produced as the laser light is absorbed will: 1) weaken the covalent bonding, particularly the bond bending forces, and render the material much more plastic than normal; and 2) detrap point defect clusters which are normally bound coulombically so that the defects will be much more mobile. In the presence of a dense (>10^{21}/cm^3) electron-hole plasma which may be produced by the intense irradiation during pulsed laser annealing, the covalent bonding of tetrahedral semiconductors like Si will be severely weakened because a significant fraction of the bond charge has been excited across the gap into antibonding or plane wave states. The crystal structure may even become unstable and undergo a phase transition if the bond charge is reduced beyond a critical point which would depend on lattice temperature, T_L. If T_L = 0 K, this transition would be second order and should occur where the transverse acoustic modes go to zero frequency. This transition has previously been estimated /35/ from the temperature dependence of the gap to occur at a density of 8×10^{21}/cm^3. At finite T_L the transition should be weakly first order because the atomic motion will increase as the bond bending forces are weakend until the structure becomes unstable. In this state the material would not be able to support a shear and so should be called a liquid, but it

would be very distinct from molten Si. The energy from the laser would be concentrated in the electronic excitation and not in kinetic energy of the atoms, as in the thermal case. Recrystallization could occur without the destructive effects of severe thermal gradients when the material passes back through this phase transition to the covalently bonded phase. It would be necessary that the plasma density fall sufficiently slowly that the bonding forces remain weak for a sufficient time so that the atoms could regain long range order in the normal tetrahedral phase. If the density were to fall too rapidly, one should expect the resultant material to be amorphous and tetrahedrally bonded. (A completely reversible crystalline to tetrahedrally bonded amorphous transition has been observed /36,37/ to occur under certain laser irradiation conditions. The author feels the transition from the dense plasma phase is a much more likely cause of this amorphous material than is the normal molten phase. Reasons for this belief are that molten Si is more than fourfold coordinated and 15 % more dense than crystalline Si, that glasses quenched from a melt retain the structure of the melt, that no one has succeeded in making amorphous Si by splat-cooling, and that rapid quenching from a high temperature would be expected to damage the material). However, recrystallization of ion-implanted material should also be possible without passing through this phase transition if the atoms are reordered by the optically induced gliding of dislocations out of the damaged zone, as has been observed in crystalline semiconductors /38/. Monemar, et al., /38/ proved that when this process occurred in GaAlAs double heterostructures, it was nonthermal. Porter, et al., /39,40/ observed laser annealing of heavily dislocated Si and showed that the effect could not be the result of epitaxial regrowth from any phase because there was no epitaxial relation between the annealed material and the Si on any of its five sides. (The annealed material was free of dislocations and completely surrounded be heavily dislocated material. By definition, epitaxial growth on a dislocated surface produces dislocated overgrowth). This observation could only be explained by rapid glide or climb of the dislocations during the annealing event. The observations by Narayan /41/ of the removal of dislocation loops and P precipitates from Si during pulsed laser annealing have to be explained in the same way because epitaxial regrowth would have produced material that was not single crystal above the precipitates and dislocations that remained in the phase boundary. Point defects should be removed from amorphous or damaged material at a rate much greater than normal furnace annealing due to five separate effects. First, they may be eliminated by recrystallization following the phase transition. Second, they may experience recombination enhanced diffusion within the covalent phase /42/. Third, the coulombic trapping by charged impurities, which normally reduces the net rate of vacancy (or interstitial) migration drastically /43/, will be suppressed by the flux of free carriers in the dense plasma. Fourth, point defects may be swept out of the material as the large concentration of dislocations, which are ideal sinks, glide out. Fifth, the energy of formation of fast diffusing interstitial species /44/ can be greatly reduced by the dense plasma.

We must now elaborate upon the third point. There must be at least some moderate lattice heating of the sample during pulsed laser annealing. Therefore, let us make calculation easy by assuming $kT_L = 0.05$ eV, $T_L = 307$ C, where T_L is, again, the lattice temperature - as distinguished from the temperature of the plasma, T_e, which would be much greater. It would be very difficult to attain an accurate phonon spectrum for the material during pulsed laser annealing, but we may assume that the lattice attempt frequency, ν, is given by

$$h\nu = kT_L, \qquad (1)$$

which implies that

$$\nu = 1.2 \times 10^{13}/s \qquad (2)$$

We also do not know empirical values for the enthalpies of migration of vacancies in amorphous Si, so we approximate them with values determined for crystalline Si.

Thus, we use Watkins's measured /45/ values for the activation energy for single vacancy migration for the neutral, V^X, and double acceptor, V^{-2}, ionization states, which were noted before,

$$\Delta H_m(V^{-2}) = 0.18 \text{ eV}, \quad (3)$$

and

$$\Delta H_m(V^X) = 0.33 \text{ eV}, \quad (4)$$

and the standard expression for vacancy diffusivity in the diamond lattice

$$D(T_L) = 0.0625\, a^2\, \nu \exp(\Delta S_m/k) \exp(-\Delta H_m/kT_L), \quad (5)$$

where a is the lattice constant of Si, a = 0.543 nm, and the factor 0.0625 results from a combination of geometrical and correlation of hoping direction considerations. ΔS_m is the entropy of vacancy migration. This has been estimated from purely geometrical grounds by Swalin /46/. That estimate has been checked against accurate data by Van Vechten and Thurmond /43/. It is

$$\Delta S_m = 4.1\ k \quad (6)$$

The result is

$$D(V^{-2}, T=307\ C) = 3.7 \times 10^{-3} \text{ cm}^2/\text{s}, \quad (7)$$

and

$$D(V^X, T=307\ C) = 1.8 \times 10^{-4} \text{ cm}^2/\text{s}, \quad (8)$$

Therefore, if the isolated vacancies migrate without trapping for τ = 200 ns., there $d = [D\tau]^{1/2}$ will be

$$d(V^{-2}, 307\ C, 200 \text{ ns.}) = 270 \text{ nm}, \quad (9)$$

and

$$d(V^X, 307\ C, 200 \text{ ns.}) = 60 \text{ nm}. \quad (10)$$

If the depth of damage is 50 to 100 nm, the vacancies would have roughly sufficient time to migrate all the way to the free surface. Note that a) as the V^{-2}'s migrate out, more V^X's are ionized to maintain the electronic equilibrium; and b) the vacancies need not migrate all the way to the free surface - they can, and apparently do, /47/ collect in voids, which may take longer, and more energy, to remove. Note that we have assumed that the untrapped migration of the vacancies occurs only as long as the high reflectivity was observed /10/ at 630 nm at the annealing threshold, which corresponds to an electron density of order 10^{21} cm^{-3}. It might well be that it actually persists significantly longer.

Next, there is the question of the diffusivity of the impurities. Because we are assuming there is no effective binding between the vacancies and impurities, we should expect, at least in first approximation, that

$$D(I) = [V^{-2}]\, D(V^{-2}) + [V^-]\, D(V^-) + [V^X]\, D(V^X) + [V^+]\, D(V^+) \quad (11)$$

i.e., that the diffusivity of atoms that migrate by vacancy motion is just the vacancy diffusivity times the probability of exchange with a vacancy, which in this case is taken to be just the atom fraction of vacancies. As the initial vacancy concentration, the excess volume is several percent /48/ the impurity diffusivity, D(I), will be several percent of the vacancy diffusivities just calculated until the excess volume is expelled. Therefore, we should estimate

$$D(I, 307\ C) = 2 \times 10^{-6} \text{ cm}^2/\text{s} \quad (12)$$

as long as the vacancies are at this high concentration. Wang, et al., /21/ who report some of the largest impurity redistributions measured, fir their data by assuming $D(I) = 2.4 \times 10^{-6}$ cm^2/s for 180 ns.

We now address the problem of surface segregation of ion implanted impurities. Impurity segregation will occur across any phase boundary where the free energy of the impurity is less on one side than on the other. The magnitude of this effect will depend upon this driving force, the diffusivity of the impurity and the time available for diffusion. If the material were melting, one would expect impurities with similar molten phase diffusivities to be segregated in proportion to their normal molten-solid segregation coefficients. (Note that with a diffusivity of 1×10^{-4} cm^2 a period, t, of only 160 ns would give $(2Dt)^{1/2}$ = 100 nm, the entire thickness of a typical damaged layer). This is not observed /21,22/ Group VI impurities are very little segregated by pulsed laser

annealing even when they are very much segregated by the normal melting transition. The Group V impurity Bi is segregated less than would be predicted from the melting hypothesis. Cu, Zn, Fe and the Group III elements Ga and In are substantially segregated by pulsed laser annealing. If the plasma remains dense long enough for the vacancies and other host lattice defects to diffuse out, then there are very few vacancies available to diffuse impurities when the plasma-normal phase boundary sweeps to the surface. Group IV, V, and Vi impurities require a vacancy (or other host lattice defect) to diffuse and therefore should be expected to segregate very little during plasma annealing. Cu, Zn, and Fe normally diffuse as interstitials without any host lattice defect and therefore should remain mobile as the plasma boundary sweeps to the surface. Therefore, such impurities should segregate providing there is a driving force.. The fact that these impurities are present above the saturation limit of the normal Si assures such a driving force. There is also the observation that there is a strong correlation between the valence of an impurity and the energy (excess heat of mixing) required to create it in a semiconductor host. The most soluble interstitials have valence 1 (Li, Na, K, Cu, Ag, and Au); elements with valence 2 are observed as interstitials but with lower solubility (Be, Mg, and Zn). In irradiated Si one finds /45/ as many Group III interstitials as there are Si vacancies that have been introduced; the Si self interstitial has never been observed and only the smallest Group IV element, C, has been observed /49/, only in irradiated Si. Group V interstitials have never been observed /45/ in irradiated or non-irradiated Si. A simple but quantitative explanation /50,51/ of this observation has been based on electronic structure and the requirement to orthogonalize the wave functions of the interstitial against those of the host lattice. In the presence of a dense electron-hole plasma the energy of formation of a Group III interstitial should be reduced by the ionization of both host and impurity atoms that interstitial diffusion could become significant in plasma phase. This would then explain the much greater degree of segregation that is found with Group III impurities than with Group V impurities having similar molten phase diffusivities and segregation coefficients.

Now let us consider the diffusion of the plasma itself. The dense plasma, which is essentially a super heated electron - hole drop, will expand into the substrate. The ambipolar diffusivity of carriers in Si at room temperature /52/ is 18 cm^2/s. Under pulsed laser annealing conditions this diffusivity would probably be greater due to an elevated value of the effective carrier temperature, T_e (which must be distinguished from the effective lattice temperature, T_L) and due to increase in phonon scattering time which we have argued must occur in order to account for the observed persistence of the plasma for several hundred nanoseconds. Indeed an increase in ambipolar diffusivity in Ge subjected to picosecond laser irradiation has been determined from direct observation of carrier diffusion /53/. It was found to be about three times the normal room temperature value. Yoffa's /26/ estimated diffusivity of a homogeneous plasma subject to irradiation typical of nanosecond laser annealing in Si is between 100 and 500 cm^2/s. Because the carrier scattering time and diffusivity is strongly dependent on plasma density and temperature, the diffusivity will vary with position and with time as the plasma expands. An accurate description of the process will have to consider the effect of the moving boundary of the dense plasma beyond which carriers will rapidly loose their energy to the lattice. Boundary effects will thus limit the effective plasma diffusivity. However, for the purpose of making a crude estimate of the magnitude of the plasma expansion, let us assume a constant effective diffusivity

$$D_P = 100 \text{ cm}^2/\text{s} \qquad (13)$$

With these simplifing approximations, we predict that the plasma expands into the substrate a distance

$$d = [D_P \tau]^{1/2} = 39 \text{ } \mu\text{m} \qquad (14)$$

for a time τ = 150 ns. This would be an expansion of a factor of order 390 from the volume in which the 533 nm light was obserbed. When and where the plasma becomes underdense, it will no longer be metastable and will dump its energy into the lattice, but this energy will be dumped into a very large volume. Whereas a typical E^a = 0.25 J/cm^2 of absorbed energy (0.5 J/cm^2 incident) would be sufficient to melt 0.15 μm of (crystalline) Si if there were no diffusion, it cannot hear 39 μm more than 40° C.

Recall that it is hypothesized that the single shot annealing threshold is determined by the requirement that the plasma be made sufficiently dense that it can remain above critical density, n_{min}, although it is expanding rapidly, long enough for the vacancies to migrate out. Empirically this time /16/ seems to be about 150 ns, which was consistent with our rough estimate of the vacancy diffusion. Consider now the density of the plasma that might be produced by the initial laser pulse. In figures 1 and 2 we plot the variation with time of the surface energy density, E(t), for laser pulses with a triangular pulse shape (in time) with full widths, 2τ = 30 ns and 30 ps respectively, a constant and

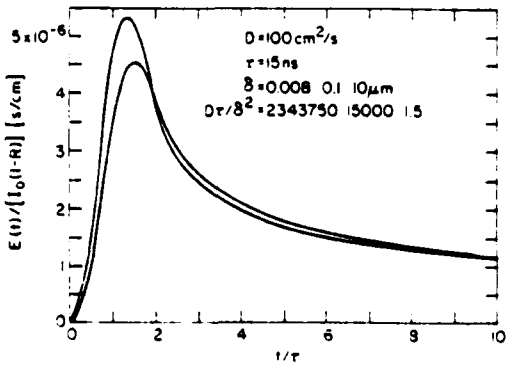


Fig. 1. Surface energy density as a function of time for a triangular laser pulse having half width τ = 15 ns for a diffusivity of 100 cm^2/s and for optical absorption depths of 8 nm, 100 nm, and 10 μm. This figure is generously supplied by R. A. Ghez.

uniform energy diffusivity of 100 cm^2/s and for optical absorption depths of 8 nm (for 266 nm light), 100 nm (for 533 nm light), and 10 μm (for 1.06 μm light). E(t) is ploted in reduced units so that values may be deduced for any value of the peak incident intensity, I_0. If the we take the value, E^a = 0.25 J/cm^2, of the absorbed (0.5 J/cm^2 incident) energy per unit area as the one shot annealing threshold /19,28/ for 533 nm green light, we would have a maximal energy density of

$$E_{max}(0.25\ J/cm^2,\ 30\ ns,\ 533\ nm) = 86\ J/cm^3 = 5.5 \times 10^{23}\ eV/cm^3. \quad (15)$$

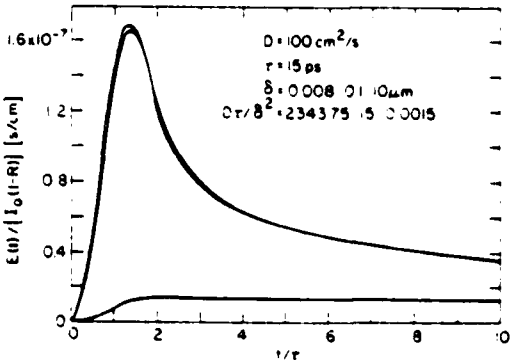


Fig. 2. Surface energy density as a function of time for a triangular laser pulse having a half width τ = 15 ps for a diffusivity of 100 cm^2/s and for optical absorption depths of 8 nm, 100 nm, and 10 μm. This figure is generously supplied by R. A. Ghez.

If we hypothesize that none of this energy is transferred from the electronic system to the lattice, but take account of the expansion of the plasma during the pulse, we have about 5×10^{23}/cm^3 excited electrons and holes at the maximum, which occurs shortly after the laser's peak, and about 1.2×10^{23}/cm^3 after 150 ns. With the 30 ps pulse, the maximum excited electron concentration would be about 1.7×10^{22}/cm^3 for 0.25 J/cm^2 and the same absorption and diffusivity. However, after 150 ns, energy and carrier density would be essentially the same as for the 30 ns pulse of the same energy. Therefore, we conclude that for this threshold pulse energy, the 30 ns pulse would not drive the material through the weakly first order phase transition, but the 30 ps pulse would. It would take an estimated 3.6 J/cm^2 to produce the phase transition to the electronically induced liquid state with a 30 ns laser pulse. However, the 30 ns pulse would drive the carrier concentration beyond 20 times the intrinsic value at the melting point and keep it more than 5 times that value

for the 150 ns required for vacancy diffusion. This level is well above that required for optically induced dislocation glide /38/, about $2\times10^{19}/cm^3$, so that the material would be rather plastic. If the effective diffusivity were equal to the handbook value, 18 cm^2/s, the carrier density would be higher and the energy requirement lower by a factor of 2.4. (It should be noted that most annealing is done with incident absorbed energies considerably higher than the 0.25 J/cm^2 threshold so that one should expect often to see the effects of the plasma fluid phase, especially at the "hot spots" of an inhomogeous laser beam).

Are these threshold concentrations reasonably consistent with the properties ascribed to them? Because it is claimed that the increased surface reflectivity results from the plasma and not from a molten layer $\hbar\omega_p$ must exceed the energy of the probing photon for the observed period. For the commonly used He-Ne laser the photon energy is 1.97 eV. In the standard expression for ω_p,

$$\omega_p = [4\pi ne^2 / \varepsilon m^*]^{1/2} \qquad (16)$$

it is not obvious what value of the carrier effective mass, m^*, or of the effective dielectric constant, ε, ought to be assumed under PLA conditions (amorphous material and a very large carrier density). However, let us estimate $\varepsilon = 12$ and $m^* = 0.2\ m_e$, where m_e is the free electron mass. Then the concentration of $1.2\times10^{20}/cm^3$ estimated to occur 150 ns after the onset of a 0.25 J/cm^2 (absorbed) 30 ns 533 nm threshold pulse would have $\hbar\omega_p = 2.6$ eV, sufficient to give the observed high reflectivity. As this is also 50 times the maximal phonon energy, plasmon-phonon coupling would also be weak /24,25,27/. According to Yoffa's estimate /26/, the intravalley phonon scattering via the deformation potential channel would be screened down by a factor of 23 at this point.

Murakami, et al. /13/ showed that annealing with a mode-locked laser (which emitted 30 ps pulses at 10 ns intervals for 150 ns) required an energy density just great enough that the reflectivity of a He-Ne laser was barely increased. This implies that the critical value of the plasma density, n_{min}, must be 25 % or so below that which would make $\hbar\omega_p = 1.97$ eV, the He-Ne photon energy. This reasoning implies that $n_{min} \approx 5\times10^{19}/cm^3$, which is twice the excited carrier concentration at the melting point.

Note that because the mode-locked laser puts in energy continuously throughout the duration of the annealing event, the maximal carrier concentration does not need to be as great as in the case of annealing with Q-switched lasers, which put all the energy in at the beginning of the event and then allow it to diffuse away. This leads to the predicition that better quality material will be produced with less power consumed using mode-locked rather than Q-switched lasers. It seems that any model which maintains the criterion for good annealing is that a sufficicient temperature or degree of melting be obtained must make the opposite prediction. This is because more heat will diffuse out of the zone in the 150 ns of a mode-locked laser pulse train than in the 10 ns or so of a Q-switched laser pulse. It seems that the prediction of the PA model is correct. /54,55/

The fact that the reflectivity decreases back to a normal Si value only rather slowly, is explained in the plasma hypothesis by the fact that the plasma edge is not a sharp step but is broadened by interaction between the plasmons and the single particle excitations, which is strong because $\hbar\omega_p$ exceeds the band gap, E_{cv}. Thus, as the plasma expands and $\hbar\omega_p$ decreases through the energy of the probing photons, the reflectivity coefficient slides down a gentle slope.

The fact that absorbed energy threshold required for annealing with 533 and 266 nm light does not change measurably /15/ between 30 pc. and 30 ns pulse durations is explained in the PA model by the consideration that the requirement for completed annealing, as opposed to the threshold for a transient reflectivity increase, is a sufficiently dense initial plasma. Once one is over the threshold

for metastability of the plasma, a sufficient energy must be put in to create this critical density and it does not matter what the time interval is so long as it is short comparated with the 150 ns, or so, that the plasma must persist to get the excess volume out. One should expect to see somewhat of a decrease with shorter pulses. Because the power is greater, less energy is wasted to the lattice before the plasma becomes metastable, but this is not a large factor. Suppose that metastability was reached only at the peak of the pulse, then some portion of the first half of the pulse would be wasted as heat. Auston, et al. /16/ found that single shot Q-switch laser annealing requires that the energy absorbed be twice that required to cause a few ns increase in the reflectivity of 630 nm He-Ne light. (We have shown above that this reflectivity implies a carrier density slightly greater than that required for metastability). Therefore, we may conservatively estimate that the reduction in absorbed energy density between 30 ns and 30 ps pulses of light that will be absorbed on a scale of order the layer thickness or less should not be greater than some portion of one quarter the energy of the 30 ns pulse. (Note that the melting model predicts the change should be a factor of 6 or more due to the diffusion of heat that occurs in 30 ns. Any theory in which the maximal surface temperature is the determining parameter will also predict a far lower threshold for 30 ps pulses than for 30 ns pulses because that maximal temperature will be larger by a factor of 30 in the case of the ps pulse).

Acknowledgements.- The author is very grateful to D. Hoonhout and F. W. Saris of the F. O. M. Institut voor Atoom-en Molecuulfysica in Amsterdam, with whom he has collaborated for much of this work. He is also grateful to R. A. Ghez and G. J. Lasher who have supplied computer programs and aid in modelling. He has enjoyed stimulating discussions with J. M. Gibson, S. Kirkpatrick, R. Rosenberg, R. Tsu, J. von Gutfeld, and E. J. Yoffa. Gibson, Tsu, and von Gutfeld have also been very hepful in supplying experimental observations.

/1/ Khaibullin, I. B., Shtyrkov, E. I., Zaripov, M. M., Zaripov, Bayazitov, R. M., and Galjautdinov, M. F., Rad. Eff. 36 (1978) 225 and references therein.

/2/ Kachurin, G. A., Bogatyriov, V. A., Romanov, S. I., and Smirnov, L. S., in Ion Implantation in Semiconductors 1976, edited by F. Chernow, J. A. Borders, and D. K. Brice (Plenum, New York) 1977 p. 445.

/3/ Dvurechensky, A. V., Kachurin, G. A., Mustafin, T. N., and Smirnov, L. S., in Laser-Solid Interactions and Laser Processing-1978, AIP Conf. Proc. No. 50, edited by S. D. Ferris, H. J. Leamy, and J. M. Poate (American Institute of Physics, New-York) 1979 p. 245

/4/ Grinberg, A. A., Nekhtiev, R. F., Ryvkin, S. M., Salmanov, V. M., and Yaroshetskii, I. D., Sov. Phys. Solid State 9 (1967) 1085.

/5/ Bertolotti, M., Stagni, L., Vitali, G., and Nuzii, L., J. Appl. Phys. 42 (1971) 5893.

/6/ Darby, F. W., and Paek, U.-C., IEEE J. Quant. Electron. QE 8 (1972) 106.

/7/ Giuliani, J. F., and Marquardt, C. L., J. Appl. Phys. 43 (1975) 4993.

/8/ Van Vechten, J. A., Tsu, R., Saris, F. W., and Hoonhout, D., Phys. Lett. 74A (1979) 417.

/9/ Tsu, R., Baglin, J. E., Tan, T., and von Gutfeld, R. J., Electrochem. Soc. Extend. Abstracts 79-2, (1979).

/10/ Sai-Halasz, G. A., to be published.

/11/ Ryvkin, S. M., Salmanov, V. M., and Yaroshetskii, I. D., Sov. Phys. - Solid State 10 (1968) 807.

/13/ Murakami, K., Kawabe, M., Gamo, K., Namba, S., and Aoyagi, Y., Phys. Lett. 70A (1979) 332.

/14/ Murakami, K., Gamo, K., Namba, S., Kawabe, M., and Aoyagi, Y., Appl. Phys. Lett. 35, to be published (Sept. 1979).

/15/ Hodgson, R. T., Tsu, R., Van Vechten, J. A., and Yoffa, E. J., Bull. Am. Phys. Soc. 24 (1979) 315.

/16/ Auston, D. H., Surko, C. M., Venkatesan, T. N. C., Slusher, R. E., and Goloychenko, Appl. Phys. Lett. 33 (1978) 437.

/17/ Auston, D. H., Golovchenko, J. A., Simons, A. L., Surko, C. M., and Venkatesan, T. N. C., Appl. Phys. Lett. 34 (1979) 777.

/18/ Birnbaum, M., and Stocker, T. L., J. Appl. Phys. 39 (1968) 6032.

/19/ Surko, C. M., Simons, A. I., Auston, D. H., Golovchenko, J. A., and Slusher, R. E., Appl. Phys. Lett. 34 (1979) 635.

/20/ Shvarev, K. M., Baum, B. A., and Geld, P. V., Sov. Phys. - Solid State 16 (1975) 2111.

/21/ Hoonhout, D., and Saris, F. W., Phys. Lett. 74A (1979) 253.

/22/ Van Vechten, J. A., Tsu, R., and Saris, F. W., Phys. Lett. 74A (1979) 422.

/23/ Matthews, M. D., and Ashby; S. J., Phil. Mag. 27 (1973) 1319 See also Germain, P., Squelard, S., and Bourgoin, J. C., J. Non-Cryst. Solids 23 (1977) 93.

/24/ Van Vechten, J. A., Yoffa, E. J., Ghez, R. A., and Tsu, R., 37th Annual Device Research Conference, Boulder, Colorado, 26 June 1979 Abstract TA1; and to published.

/25/ Yoffa, E. J., and Van Vechten, J. A., Bull. Am. Phys. Soc. 24 (1979) 315.

/26/ Yoffa, E. J., Phys.Rev.B.21 (1980).(TBP)

/27/ Tsu, R., and Jha, S. S., to be published

/28/ Tsu, R., Baglin, J. E., Tan, T. Y., Tsai, M. Y., Park, K. C., and Hodgson, R. T., in Laser-Solids Interactions and Laser Processing - 1978 (AIP Conf. Proc. 50) edited by S. D. Ferris, H. J. Leamy, and J. M. Poate, p. 344.

/29/ Shanks, H. R., Maycock, P. D., Sidles, P. H., and Danielson, C. G., Phys. Rev. 130 (1963) 1743.

/30/ Glassbrenner, C. J., and Slack, G. A., Phys. Rev. 134A (1964) 1058.

/31/ Yoffa, E.J., Laser and Electron Beam Processing of Materials Symposium, Boston, Mass., 27 November 1979, abstract no. 1-4 ; Apol. Phys. Lett. 36 (1980) 37.

/32/ Wang, J. C., Wood, R. F., and Pronko, P. P., Appl. Phys. Lett. 33 (1978) 455.

/33/ Baeri, P., Campisano, S. U., Foti, G., and Rimini, E., J. Appl. Phys. 50 (1979) 788.

/34/ Schultz, J. C., and Collins, R. J., Appl. Phys. Lett. 34 (1979) 363.

/35/ Heine, V., and Van Vechten, J. A., Phys. Rev. B 13, (1976) 1622.

/36/ Tsu, R., Hodgson, R. T., Tan, T. Y., and Baglin, J. E., Phys. Rev. Lett. 42 (1979) 1356.

/37/ Liu, P. L., Yen, R., Bloembergen, N., and Hodgson, R. T., Appl. Phys. Lett. 34 (1979) 864.

/38/ Nonemar, B., Potemski, R. M., Small, M. B., Van Vechten, J. A., and Woolhouse G. R. Phys. Rev. Lett. 41 (1978) 260. Note that amorphous, and liquid, solids are often described as crystals with a large dislocation density; see, e.g., S. F. Edwards and M. Warner, Phil. Mag. A 40 (1979) 257.

/39/ Porter, W., Parker, D. L., Richardson, T. W., and Swenson, E. J., Appl. Phys. Lett. 33 (1978) 886.

/40/ Porter, W., Parker, D. L., and Richardson, T. W., in Laser-Solid Interactions and Laser Processing-1978 (op. cit. Ref. 28) p. 453.

/41/ Narayan, J., Appl. Phys. Lett. 34 (1979) 312.

/42/ Troxell, J. R., Chatterjee, A. P., Watking, G. D., and Kimerling, L. C., Phys. Rev. B 19 (1979) 5336 and references therein.

/43/ Van Vechten, J. A., and Thurmond, C. D., Phys. Rev. B 14 (1976) 3551 and references therein.

/44/ Van Vechten, J. A., Inst. Phys. Conf. Ser. No. 31 (1977) 441.

/45/ Watkings, G. D., Inst. Phys. Conf. Ser. No. 31 (1975) 1.

/46/ Swalin, R. A., J. Phys. Chem. Solids 18 (1961) 290.

/47/ Andrew, R., and Lovato, M., J. Appl. Phys. 50 (1979) 1142.

/48/ Knights, J. C., AIP Conf. Proc. 31 (1976) 296.

/49/ Brower, K. L., Phys. Rev. B 14 (1976) 872.

/50/ Van Vechten, J. A., in Radiation Effects in Semiconductors 1976, edited by N. B. Urli and Corbett, J. W., (Inst. of Physics, London, 1977). Inst. Phys. Conf. Ser. No. 31, pp. 441.

/51/ Van Vechten, J. A., in Semiconductor Handbook, Vol. 3 edited by S. P. Keller (North-Holland, Amsterdam) 1980 pp. ?

/52/ Wolf, H., Semiconductors (John Wiley, New-York) 1971 p. 290.

/53/ Auston, D. H., and Shank, C. V., Phys. Rev. Lett. 32 (1974) 1120.

/54/ Saris, F. W., private communication.

/55/ Rimini, E., patents applied.

Appendix C

Volume 74A, number 6 PHYSICS LETTERS 10 December 1979

NONTHERMAL PULSED LASER ANNEALING OF Si: PLASMA ANNEALING

J.A. Van VECHTEN [1] and R. TSU
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

and

F.W. SARIS
FOM Instituut voor Atoom- en Molecuulfysica, Amsterdam, The Netherlands

Received 3 September 1979

In the accompanying paper we have given evidence that pulsed laser annealing of Si does not involve normal thermal melting and recrystallization. Here we argue the importance of the electron–hole plasma produced by the laser to the annealing process.

In the accompanying [1] paper reasons have been given to doubt that the pulsed laser annealing of Si involves normal thermal melting. Here we try to answer two questions: (1) how can the material escape melting; and (2) what is the role of the electron–hole plasma produced by absorption of the laser light.

How can the material escape melting. The energy required to heat crystalline Si from room temperature to the melting point and to melt it is [2] 7.3×10^3 J/cm^3. Due to the heat of recrystallization (which has not yet been determined experimentally for Si), the energy required to heat and melt amorphous Si is somewhat, perhaps a third, less [3]. Therefore, the energy required to melt a typical 100 nm layer of ion implanted Si would not exceed 0.073 J cm^{-2} if there were no diffusion of energy to the substrate. As the normal incidence reflectivity rises from about 35% at the beginning of the pulse to about 60% during the observed high reflectivity period [4–6] the energy absorbed is about half that incident. Thus the incident energy density threshold required to melt a 100 nm layer with no diffusion could not exceed 0.15 J/cm^2 and would be more like 0.10 J cm^{-2} for an amorphous ion implanted surface.

[1] Supported in part by the Air Force Office of Scientific Research under Contract No. F49620-79-C-0077.

The observed incident energy threshold for single shot complete annealing of a 100 nm amorphous layer with either 30 ns or 30 ps pulses is about 0.5 J/cm^2 for 533 nm light [7,8] which is absorbed in the same 100 nm or about 0.25 J/cm^2 for 266 nm light [8] which is absorbed in the first 8 nm. Therefore, at this threshold several times more energy is absorbed than would be required to melt the amorphous (damaged) layer if it were all delivered as heat in that layer and none diffused out. The question whether the surface layer melts, gets hot without melting or is warmed only moderately turns on the rates at which (a) energy is transferred from the electrons and holes created by the laser to the lattice as heat, (b) this plasma of carriers expands into the substrate, and (c) the heat delivered to the lattice diffuses into the substrate.

The energy of a 533 nm photon exceeds the energy gap of Si by more than 1 eV, this energy would appear as kinetic energy of the carriers excited across the gap. Free carrier absorption would leave the carrier with more than 2.3 eV of kinetic energy. Even if a phonon emission time of 10^{-13} s is assumed [9] these carriers would still diffuse more than 200 nm before giving up 1 eV to the lattice. Thus alone would double or triple the energy required to melt 100 nm with 533 or 266 nm light. The thermal diffusivity of Si under

normal furnace conditions (in the dark) varies from 0.86 cm^2/s at room temperature to 0.117 cm^2/s at the melting point [10,11]. If we take $D = 0.117$ and $\tau = 30$ ns, a typical laser pulse duration, and if we make a rough estimate of the extent of thermal diffusion, we find $(D\tau)^{1/2} = 600$ nm or six times the depth of the layer to be annealed. (For a more detailed account of these two effects see the recent study of Yoffa [12].) In fact the diffusion of the heat that has been delivered to the lattice will be substantially greater than this estimate because the laser pulse produces a carrier concentration which, by all estimates, is much larger than the thermal intrinsic value (2×10^{19} cm^{-3} at the melting point and less at lower temperatures) and because 40% of the thermal diffusivity at the melting point is due to the intrinsic free carriers [11]. If we make the conservative estimate [12] that the carrier concentration during the laser pulse is 2×10^{20} cm^{-3}, then the diffusivity at the melting point would be 0.54 cm^2/s and $(D\tau)^{1/2} = 1.3$ μm. Combining this consideration with that of the diffusion of the carriers even with a 10^{-13} s phonon emission time, we can conclude that the damaged layer will not be melted back to the substrate at the observed incident energy density thresholds of 0.25 J/cm^2 for 266 nm or 0.5 J/cm^2 for 533 nm light. (Some published computer simulations [7,13–15] of the heating of the material during pulsed laser annealing have concluded that the annealing threshold is the level at which melting does extend to the damage interface. These neglect the effects of carrier diffusion as phonons are being emitted and the enhancement of thermal diffusivity due to the excess carrier concentration produced by the intense laser light.)

We interpret the often reported rise in surface reflectivity as the result of an electron–hole plasma of sufficient density that $\hbar\omega_p$ exceeds the energy of the photons used to determine the reflectivity. This is the conventional interpretation of the surface reflectivity rise [16–18], for example that given for the Q-switching effect of Si mirrors, but contrary to the recent suggestion of Auston et al. [4] that the effect might be due to melting of the surface. In the accompanying paper [1] we showed that the melting hypothesis does not account for the manner in which the reflectivity rises and falls and that there is no close agreement between the measured reflectivity during the period that it is increased and the reflectivity of normal molten Si. Therefore, we contend that the observation of the duration of the period of enhanced surface reflectivity, from a few ns to as much as several hundred ns, indicates that the transfer of energy from the dense plasma to the lattice requires a comparable time, much longer than 10^{-13} s, under pulsed laser annealing conditions. If we take the phonon emission time for the dense plasma to be several orders of magnitude longer than 1×10^{-13} s, then we must conclude that the surface is not even warmed more than a few 100°C.

In separate publications [19,20] we explain why this rate should be so much slower than that extrapolated from measurements at lower levels of irradiation. These explanations involve: (a) the decoupling of the lattice from the plasmons of the excited carriers when $\hbar\omega_p$ becomes much larger than all phonon modes (when the excited carrier density exceeds 10^{19} cm^{-3}); (b) the screening of the deformation potential coupling between hot carriers and phonons (an effect which Yoffa has treated in some detail [21]); (c) an argument that under normal conditions the coupling of the lattice to the carriers via the plasmons is more important than that via the deformation potential, particularly for intervalley scattering with the emission of longitudinal phonons; and (d) a calculation [20] showing that in Si the only optic phonons coupling to plasmons have (100) wave vectors so that plasmons with wave vectors in other directions will be long lived when $\hbar\omega_p$ is much larger than acoustic branch energies.

Plasma annealing. How can the material be annealed if it does not melt but does have a dense electron–hole plasma excited within it? In order for the material to be annealed two things must occur. The crystalline order must be reestablished and the large supersaturation of lattice defects, or equivalently the excess volume [22] (about 9%) of the damaged layer with respect to the crystal, must be diffused out.

There are two ways the crystalline order may be reestablished without melting. If more than a critical number of electrons, which has been estimated to be 8×10^{21} cm^{-3} from considerations of Si band structure and temperature dependence of its band gaps [23–25], are excited out of the bonding states of the valence band into the antibonding states of the conduction band, then a second order phase transition will occur. The bond charges will be so depleted that they

will no longer be able the stabilize the TA phonon modes [26]; the crystal will no longer resist shearing stresses but will become fluid. This fluid is distinct from the normal molten phase of Si because the latter is the result of a strictly first order phase transition that is driven by the violent motion of the atoms at high temperature. In the plasma case we maintain that the energy is still retained in the electronic system so that the atoms are not in violent motion. As the plasma becomes less dense due to expansion and due to transfer of energy to the lattice, the material will pass back through the second order phase transition at 8×10^{21} cm^{-3} and the covalent bonding will gradually reappear. If this process is not too fast, the material will recrystallize as epitaxial single crystal because this is the state of lowest energy. (We believe that the crystalline to (fourfold coordinated) amorphous transition that has been observed [27,28] with 30 ps laser pulses of low energy results from passing through this second order transition so rapidly that the atoms were not able to establish long ranged order in the covalent phase before they lost all mobility. The hypothesis that the material was thermally melted and quenched like a glass suffers from the fact that molten Si has a nontetrahedral structure with a coordination greater than four (6 or more depending on temperature [29]) and glasses retain the structure of the molten phase [30]). Even if this critical carrier density for the second order phase transition is not reached, crystalline order may be regained by sufficiently weakening the covalent bond bending forces that the atoms will rearrange themselves in response to the stresses imposed by the crystalline substrate or any other source. This process is analogous to that of the optically induced gliding of dislocations in crystalline material that has been observed at much lower power level [31]. (One may view the reordering as the gliding of a large number of dislocations out of the damaged region.)

We believe that the rate limiting step in the annealing process is the removal of the excess lattice defects that were introduced by ion implantation and the concomitant excess volume. As an amorphous damaged layer is less dense typically by about 9% than the crystalline material [22] and has essentially the same nearest neighbor distance [32], we view the excess volume as a 9% concentration of trapped vacancies.

Now, vacancies are known [33] to be mobile in Si at 80 K. The reason they do not migrate out at room temperature is that they have both donor and acceptor charge states and become bound coulombically to other charged centers in the lattice. In the presence of a dense electron–hole plasma, these vacancies should diffuse with negligible trapping. If one simply takes the migration rates and activation energies measured at low temperatures by Watkins and extrapolates to some moderate temperature, e.g. 307°C, one finds the diffusivity of the vacancy in its double acceptor state, for example, to be 3.7×10^{-3} cm^2/s. Then if one takes $\tau = 160$ ns, a typical duration for the observed high surface reflectivity, $(D(\mathrm{V})\tau)^{1/2} = 240$ nm. This would seem to be enough to insure that most of the excess vacancies could diffuse out of a typical 100 nm thick damaged layer.

The rate of impurity diffusion by the vacancy mechanism would be just the vacancy diffusivity times the fraction of vacancies present. Assuming an initial concentration of about 9%, this would give an initial impurity diffusivity of 3.3×10^{-4} cm^2/s, which would be adequate to account for the impurity redistribution that is observed in pulsed laser annealing. (A typical example was fitted [34] assuming a constant diffusivity of 2.4×10^{-4} for 180 ns. In the plasma case the diffusivity should not be constant as the number of excess vacancies decreases during the event.) If the material undergoes the second order phase transition to the plasma fluid, then the redistribution will be increased during that period, but the concentrations of vacancies (excess volume) should be reduced by volume contraction of this fluid, so that after return to the covalently bonded phase the diffusion of impurities via the vacancy mechanism should be suppressed.

Finally we address the problem of surface segregation of ion implanted impurities. Impurity segregation will occur across any phase boundary where the free energy of the impurity is less on one side than on the other. The magnitude of this effect will depend upon this driving force, the diffusivity of the impurity and the time available for diffusion. If the material were melting, one would expect impurities with similar molten phase diffusivities to be segregated in proportion to their normal molten–solid segregation coefficients. (Note that with a diffusivity of 10^{-4} cm^2/s and a period, t, of only 160 ns would give $(2Dt)^{1/2} = 100$ nm, the entire thickness of a typical damaged layer.) This is not observed [35]. Group VI impurities are very little

segregated by pulsed laser annealing even when they are very much segregated by the normal melting transition. The group V impurity Bi is segregated less than would be predicted from the melting hypothesis. Cu, Zn, Fe and the group III elements Ga and In are substantially segregated by pulsed laser annealing. If the plasma remains dense long enough for the vacancies and other host lattice defects to diffuse out, then there are very few vacancies available to diffuse impurities when the plasma–normal phase boundary sweeps to the surface. Group IV, V, and VI impurities require a vacancy (or other host lattice defect) to diffuse and therefore should be expected to segregate very little during plasma annealing. Cu, Zn, and Fe normally diffuse as interstitials without any host lattice defect and therefore should remain mobile as the plasma boundary sweeps to the surface. Therefore, such impurities should segregate providing there is a driving force. The fact that these impurities are present above the saturation limit of the normal Si assures such a driving force. There is also the observation that there is a strong correlation between the valence of an impurity and the energy (excess heat of mixing) required to create it in a semiconductor host. The most soluble interstitials have valence 1 (Li, Na, K, Cu, Ag, and Au); elements with valence 2 are observed as interstitials but with lower solubility (Be, Mg, and Zn). In irradiated Si one finds [33] as many group III interstitials as there are Si vacancies that have been introduced; the Si self interstitial has never been observed and only the smallest group IV element, C, has been observed [36], only in irradiated Si. Group V interstitials have never been observed [33] in irradiated or nonirradiated Si. One of us (J.A.V.V.) has offered [37,38] a simple but quantitative explanation of this observation based on electronic structure and the requirement to orthogonalize the wavefunctions of the interstitial against those of the host lattice. We suggest that in the presence of a dense electron–hole plasma the energy of formation of a group III interstitial should be reduced by the ionization of both host and impurity atoms that interstitial diffusion could become significant in plasma phase. This would then explain the much greater degree of segregation that is found with group III impurities than with group V impurities having similar molten phase diffusivities and segregation coefficients.

In conclusion, we feel that the hypothesis of annealing of a relatively cool lattice in the presence of a dense electron–hole plasma, plasma annealing, offers a much more satisfactory explanation for the process which produces remarkably high quality material during pulsed laser annealing of Si than does the hypothesis of simple thermal melting with subsequent rapid quenching.

References

[1] J.A. Van Vechten, R. Tsu, F.W. Saris and D. Hoonhout, preceding paper.
[2] W.R. Runyan, Silicon semiconductor technology (McGraw-Hill, New York, 1965) pp. 213–215.
[3] D. Turnbull, private communication; see also H.S. Chen and D. Turnbull, J. Appl. Phys. 40 (1969) 4214.
[4] D.H. Auston, C.M. Surko, T.N.C. Venkatesan, R.E. Slusher and J.A. Golovchenko, Appl. Phys. Lett. 33 (1978) 437.
[5] K. Murakami, M. Kawabe, K. Gamo, S. Namba and Y. Aoyagi, Phys. Lett. 70A (1979) 332.
[6] R.T. Hodgson, R. Tsu, J.A. Van Vechten and E.J. Yoffa, Bull. Am. Phys. Soc. 24 (1979) 315.
[7] C.M. Surko, A.L. Simons, D.H. Auston, J.A. Golovchenko and R.E. Slusher, Appl. Phys. Lett. 34 (1979) 635.
[8] R. Tsu, J.E. Baglin, T.Y. Tan, M.Y. Tsai, K.C. Park and R.T. Hodgson, in: Laser–solids interactions and laser processing, eds. S.D. Ferris, H.J. Leamy and J.M. Poate, AIP Conf. Proc. 50 (1978) p. 344.
[9] N.O. Folland, Phys. Rev. B1 (1970) 1648.
[10] H.R. Shanks, P.D. Maycock, P.H. Sidles and C.G. Danielson, Phys. Rev. 130 (1963) 1743.
[11] C.J. Glassbrenner and G.A. Slack, Phys. Rev. 134A (1964) 1058.
[12] E.J. Yoffa, to be published.
[13] J.C. Wang, R.F. Wood and P.P. Pronko, Appl. Phys. Lett. 33 (1978) 455.
[14] P. Baeri, S.U. Campisano, G. Foti and E. Rimini, J. Appl. Phys. 50 (1979) 788.
[15] J.C. Schultz and R.J. Collins, Appl. Phys. Lett. 34 (1979) 363.
[16] M. Birnbaum and T.L. Stocker, J. Appl. Phys. 39 (1968) 6032.
[17] J.P. Woerdman, Philips Res. Rep. S7 (1971) 1.
[18] D.H. Auston and C.V. Shank, Phys. Rev. Lett. 32 (1974) 1120.
[19] J.A. Van Vechten, to be published.
[20] R. Tsu and S.S. Jha, to be published.
[21] E.J. Yoffa, to be published.
[22] J.C. Knights, AIP Conf. Proc. 31 (1976) p. 296.
[23] J.A. Van Vechten, in: Lattice defects in semiconductors (1974), (Institute of Physics, London, 1975) p. 212.
[24] V. Heine and J.A. Van Vechten, Phys. Rev. B13 (1976) 1622.
[25] J.A. Van Vechten, Phys. Rev. B13 (1976) 946.

[26] R.M. Matin, Phys. Rev. 186 (1969) 871.
[27] R. Tsu, R.Y. Hodgson, T.Y. Tan and J.E. Baglin, Phys. Rev. Lett. 42 (1979) 1356.
[28] P.L. Liu, R. Yen, N. Bloembergen and R.T. Hodgson, Appl. Phys. Lett. 34 (1979) 864.
[29] V.M. Glazov, S.N. Chizhevskaya and N.N. Glagoleva, Liquid semiconductors (Plenum, New York, 1969) p. 67.
[30] Cf., e.g., M. Goldstein, in: Phase transitions (1973), ed. L.E. Cross (Pergamon, New York, 1973) p. 263 and references therein.
[31] B. Monemar, R.M. Potemski, M.B. Small, J.A. Van Vechten and G.R. Woolhouse, Phys. Rev. Lett. 41 (1978) 260.
[32] S.C. Moss and J.F. Craczyk, in: Proc. Tenth Intern. Conf. Phys. semiconductors, eds. S.P. Keller, J.C. Hensel and F. Stern (U.S.A.E.C., Springfield, VA., 1970) p. 658.
[33] G.D. Watkins, in: Lattice defects in semiconductors (1974), (Institute of Physics, London, 1975) p. 1.
[34] J.C. Wang, R.F. Wood and P.P. Pronko, Appl. Phys. Letters 33 (1978) 455.
[35] D. Hoonhout and F.W. Saris, Phys. Lett. 74A (1979) 253.
[36] K.L. Brower, Phys. Rev. B14 (1976) 872.
[37] J.A. Van Vechten, in: Radiation effects in semiconductors (1976), eds. N.B. Urli and J.W. Corbett (London Conf. Ser. 31 (Institute of Physics, 1977) p. 441.
[38] J.A. Van Vechten, in: Semiconductor handbook, Vol. 3, ed. S.P. Keller (North-Holland, Amsterdam, 1979) p. 1.

RC 7927 (#33918) 8/27/79
Solid State Physics 10 pages

Defect Diffusion During Pulsed Laser Annealing*

J. A. Van Vechten
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598
U.S.A.

Abstract

In the presence of a dense ($> 10^{21}/cm^3$) electron-hole plasma which may be produced by the intense irradiation during pulsed laser annealing, the covalent bonding of tetrahedral semiconductors like Si will be severely weakened because a significant fraction of the bond charge has been excited across the gap into antibonding or plane wave states. The crystal structure may even become unstable and undergo a second order phase transition if the bond charge is reduced beyond a critical point where the transverse acoustic modes go to zero frequency. This transision has previously been estimated from the temperature dependence of the gap to occur at a density of $8x10^{21}/cm^3$. In this state the material would not be able to support a shear and so should be called a liquid, but it would be very distinct from molten Si. The energy from the laser would be concentrated in the electronic excitation and not in kinetic energy of the atoms, as in the thermal case. Recrystallization could occur without the destructive effects of severe thermal gradients when the material passes back through this second order phase transition to the covalently bonded phase. However, recrystallization of ion-implanted material should also be possible without passing through this second order phase transition if the atoms are reordered by the optically induced gliding of dislocations out of the damaged zone, as has been observed in crystalline semiconductors. Point defects should be removed from amorphous or damaged material at a rate much greater than normal furnace annealing due to five separate effects. First, they may be eliminated by recrystallization following the second order phase transition. Second, they may experience recombination enhanced diffusion within the covalent phase. Third, the coulombic trapping by charged impurities, which normally reduces the net rate of vacancy (or interstitial) migration drastically, will be suppressed by the flux of free carriers in the dense plasma. Fourth, point defects may be swept out of the material as the large concentration of dislocations, which are ideal sinks, glide out. Fifth, the energy of formation of fast diffusing interstitial species can be greatly reduced by the dense plasma. It is claimed that these effects are required to account for the very high crystallographic, morphological and electrical quality of Si produced by pulsed laser annealing.

* Supported in part by Air Force Office of Scientific Research under Contract No. F49620-79-C-0077.

In his 1977 review of the subject of pulsed laser annealing I. B. Khaibullin concluded[1], "The mechanism of laser annealing is not yet finally established. But even now one can state that in the case of the nanosecond regimes one cannot reduce it to the ordinary thermal effect. Different factors such as photoionization, impact wave, powerful light fields etc. play a significant role." In addition to the convincing arguments that pulsed laser annealing could not be a simple thermal effect and could not be simple thermal melting followed by rapid crystallization as quench which are given in that paper, several more observations and arguments lead to the same conclusion. There is, of course, the fact that long standing theories for laser damage[2-5] indicate that the material should be heavily damaged, not annealed, if all the energy of the laser pulse were converted to local heating of the material in times of order the duration of the pulse. Simple analysis[6] of the mechanical forces that would be generated by thermal expansion and by contraction upon melting in times of order the duration of these laser pulses show that a molten layer could not remain in contact with the substrate unless it balled-up to an extent completely inconsistent with the very flat surface morphology that is obtained. Thermoelastic measurements[7] made during pulsed laser annealing of Si show no evidence for the snap-back force that would have to occur if the surface melted and contracted until one increases the laser energy above the damage threshold. In addition to the observation that annealing may occur without any significant increase in surface reflectivity,[1,8] it has been noted[6] that when a rise in surface reflectivity is observed[9,10] it rises and falls in a manner expected for a dense electron-hole plasma and not that expected for molten Si. Moreover, there is no close agreement between the measured reflectivity during the high reflectivity period and the known reflectivity of molten Si. Furthermore, the redistribution of dopants has been shown[6,11,12] to be inconsistent with the hypothesis of crystallization from normal molten Si.

It is the author's opinion that it is the photoionization, the dense electron-hole plasma created by the intense laser pulse, which principally accounts for the remarkable properties of

pulsed laser annealing and that the process is indeed nonthermal in the sense that the thermal transients that do occur are neither helpful to the annealing nor particularly severe. Let us call this the "Plasma Annealing", PA, theory and process.

By far the most difficult aspect of the PA theory of pulsed laser annealing is the question why the plasma remains dense for such a long time and transfers its energy to the lattice so slowly. However, it seems that the observations that the increased surface reflectivity, which has the properties of a dense plasma and not those of a molten layer, may persist as long as 800 ns demonstrate the point experimentally. Explanations for this long lifetime and slow transfer are given in detail elsewhere.[12-15] These explanations involve: a) the decoupling of the lattice from the plasmons of the excited carriers when $\hbar\omega_p$ becomes much larger than all phonon modes[13] (when the excited carrier density exceeds $10^{19}/cm^3$); b) the screening of the deformation potential coupling between hot carriers and phonons[14]; c) an argument that under normal conditions the coupling of the lattice to the carriers via the plasmons is more important than that via the deformation potential, particularly for intervalley scattering with the emission of longitudinal phonons[13]; and d) a calculation showing that in Si the only optic phonons coupling to plasmons have (100) wave vectors so that plasmons with wave vectors in other directions will be long lived when $\hbar\omega_p$ is much larger than acoustic branch energies.[15]

However, let us note that even if there were no slowing of the energy transfer from hot carriers to the lattice, even if the phonon emission time were 10^{-13} s, one would not melt a typical 100 nm thick damaged layer back to the substrate at the observed threshold for annealing. The energy required to heat crystalline Si from room temperature to the melting point and to melt it is 7.3×10^3 J/cm^3. Due to the heat of recrystallization (which has not yet been determined experimentally for Si), the energy required to heat and melt amorphous Si is somewhat, perhaps a third, less. Therefore, the energy required to melt a 100 nm layer of ion implanted Si be about 0.05 J/cm^2 if there were no diffusion of energy to the substrate. As the normal incidence reflectivity rises from about 35 % at the beginning of the pulse to about 60

% during the observed high reflectivity period, the energy absorbed is about half that incident. Thus the incident energy density threshold required to melt a 100 nm layer with no diffusion would be about 0.10 J/cm^2 for an amorphous ion implanted surface.

The observed incident energy threshold for single shot complete annealing of a 100 nm amorphous layer with either 30 ns or 30 ps pulses is about 0.5 J/cm^2 for 533 nm light[16,17] which is absorbed in the same 100 nm or about 0.25 J/cm^2 for 266 nm light which is absorbed in the first 8 nm. The energy of a 533 nm photon exceeds the energy gap of Si by more than 1 eV; this energy would appear as kinetic energy of the carriers excited across the gap. Free carrier absorption would leave the carrier with more than 2.3 eV of kinetic energy. Auger recombination would add 1.1 eV to the kinetic energy of the third carrier in the process. Even if a phonon emission time of 1×10^{-13} s is assumed, these carriers would still diffuse more than 200 nm before giving up 1 eV to the lattice. This alone would double or triple the energy required to melt 100 nm with 533 or 266 nm light. The thermal diffusivity of Si under normal furnace conditions (in the dark) varies from 0.86 cm^2/s at room temperature to 0.117 cm^2/s at the melting point.[18,19] If we take D = 0.117 and τ = 30 ns, a typical laser pulse duration, and if we make a rough estimate of the extent of thermal diffusion, we find $(D\tau)^{1/2}$ = 600 nm or six times depth of the layer to be annealed. (For a more detailed account of these two effects see the recent study of E. J. Yoffa.[20]) In fact the diffusion of the heat that has been delivered to the lattice will be substantially greater than this estimate because the laser pulse produces a carrier concentration which, by all estimates, is much larger than the thermal intrinsic value (2×10^{19} /cm^3 at the melting point and less at lower temperatures) and because 40 % of the thermal diffusivity at the melting point is due to the intrinsic free carriers.[19] If we make the conservative estimate[14] that the carrier concentration during the laser pulse is 2×10^{20}/cm^3, then the diffusivity at the melting point would be 0.54 cm^2/s and $(D\tau)^{1/2}$ = 1.3 μm. Combining this consideration with that of the diffusion of the carriers even with a 1×10^{-13} s phonon emission time, we can conclude that the damaged layer will not be melted back to the substrate at the observed incident energy density thresholds of 0.25 J/cm^2 for 266

nm or 0.5 J/cm^2 for 533 nm light. (Some published computer simulations[16,21-23] of the heating of the material during pulsed laser annealing have concluded that the annealing threshold is the level at which melting does extend to the damage interface. These neglect the effects of carrier diffusion as phonons are being emitted and the enhancement of thermal diffusivity due to the excess carrier concentration produced by the intense laser light.)

Now let us outline the PA hypothesis. In the presence of a dense ($> 10^{21}$/cm^3) electron-hole plasma which may be produced by the intense irradiation during pulsed laser annealing the covalent bonding of tetrahedral semiconductors like Si will be severely weakened because a significant fraction of the bond charge has been excited across the gap into antibonding or plane wave states. The crystal structure may even become unstable and undergo a second order phase transition if the bond charge is reduced beyond a critical point, which has previously been estimated[24] from the temperature dependence of the gap to occur at a density of 8×10^{21}/cm^3, where the transverse acoustic modes go to zero frequency. In this state the material would not be able to support a shear and so should be called a liquid, but it would be very distinct from molten Si. The energy from the laser would be concentrated in the electronic excitation and not in kinetic energy of the atoms, as in the thermal case. Recrystallization could occur without the destructive effects of severe thermal gradients when the material passes back through this second order phase transition to the covalently bonded phase. However, recrystallization of ion-implanted material should also be possible without passing through this second order phase transition if the atoms are reordered by the optically induced gliding of dislocations out of the damaged zone, as has been observed in crystalline semiconductors.[25] Point defects should be removed from amorphous or damaged material at a rate much greater than normal furnace annealing due to five separate effects. First, they may be eliminated by recrystallization following the second order phase transition. Second, they may experience recombination enhanced diffusion within the covalent phase.[26] Third, the coulombic trapping by charged impurities, which normally reduces the net rate of vacancy (or

interstitial) migration drastically,[27] will be suppressed by the flux of free carriers in the dense plasma. Fourth, point defects may be swept out of the material as the large concentration of dislocations, which are ideal sinks, glide out. Fifth, the energy of formation of fast diffusing interstitial species[28] can be greatly reduced by the dense plasma.

We must now elaborate upon the third point. There must be at least some moderate heating of the sample during pulsed laser annealing. Therefore, let us make calculation easy by assuming $kT_l = 0.05$ eV, $T_l = 307$ C, where T^l is the lattice temperature - as distinguished from the temperature of the plasma, T_e, which would be much greater. It would be very difficult to attain an accurate phonon spectrum for the material during pulsed laser annealling, but we may assume that the lattice attempt frequency, ν, is given by

$$h\nu = kT_l, \tag{1}$$

which implies that

$$\nu = 1.2 \times 10^{13}/\text{sec}. \tag{2}$$

We also do not know empirical values for the enthalpies of migration of vacancies in amorphous Si, so we approximate them with values determined for crystalline Si. Thus, we use Watkins's measured[29] values for the activation energy for single vacancy migration for the neutral, V^x, and double acceptor, V^{-2}, ionization states, which were noted before,

$$\Delta H_m(V^{-2}) = 0.18 \text{ eV}, \tag{3}$$

and

$$\Delta H_m(V^x) = 0.33 \text{ eV}, \tag{4}$$

and the standard expression for vacancy diffusivity in the diamond lattice

$$D(T_l) = 0.0625\, a^2\, \nu \exp(\Delta S_m/k) \exp(-\Delta H_m/kT_l), \tag{5}$$

where a is the lattice constant of Si, $a = 0.543$ nm, and the factor 0.0625 results from a combination of geometrical and correlation of hoping direction considerations. ΔS_m is the entropy of vacancy migration. This has been estimated from purely geometrical grounds by

Swalin.[30] That estimate has been checked against accurate data by Van Vechten and Thurmond.[27] It is

$$\Delta S_m = 4.1 \text{ k.} \tag{6}$$

The result is

$$D(V^{-2}, T=307 \text{ C}) = 3.7 \times 10^{-3} \text{ cm}^2/\text{sec.}, \tag{7}$$

and

$$D(V^x, T=307 \text{ C}) = 1.8 \times 10^{-4} \text{ cm}^2/\text{sec.} \tag{8}$$

Therefore, if the isolated vacancies migrate without trapping for τ = 200 ns., there $d = [D\tau]^{1/2}$ will be

$$d(V^{-2}, 307 \text{ C}, 200 \text{ ns.}) = 270 \text{ nm.} \tag{9}$$

and

$$d(V^x, 307 \text{ C}, 200 \text{ ns}) = 60 \text{ nm.} \tag{10}$$

If the depth of damage is 50 to 100 nm, the vacancies would have roughly sufficient time to migrate all the way to the free surface. Note that a) as the V^{-2}'s migrate out, more V^x's are ionized to maintain the electronic equilibrium; and b) the vacancies need not migrate all the way to the free surface - they can, and apparently do,[31] collect in voids, which may take longer, and more energy, to remove. Note that we have assumed that the untrapped migration of the vacancies occurs only as long as the high reflectivity was observed[10] at 630 nm at the annealing threshold, which corresponds to an electron density of order 10^{21} cm^{-3}. It might well be that it actually persists significantly longer.

Next, there is the question of the diffusivity of the impurities. Because we are assuming there is no effective binding between the vacancies and impurities, we should expect, at least in first approximation, that

$$D(I) = [V^{-2}]\, D(V^{-2}) + [V^-]\, D(V^-) + [V^x]\, D(V^x) + [V^+] D(V^+ \tag{11}$$

i.e., that the diffusivity of atoms that migrate by vacancy motion is just the vacancy diffusivity times the probability of exchange with a vacancy, which in this case is taken to be just the

atom fraction of vacancies. As the initial vacancy concentration, the excess volume is several percent,[32] the impurity diffusivity, D(I), will be several percent of the vacancy diffusivities just calculated until the excess volume is expelled. Therefore, we should estimate

$$D(I, 307\ C) = 2 \times 10^{-4}\ cm^2/sec. \qquad (12)$$

as long as the vacancies are at this high concentration. Wang, et al.,[21] who report some of the largest impurity redistributions measured, fit their data by assuming $D(I) = 2.4 \times 10^{-4}$ $cm^2/sec.$ for 180 ns.

References

1) I. B. Khaibullin, E. I. Shtyrkov, M. M. Zaripov, R. M. Bayazitov and M. F. Galjautdinov, Radiation Effects 36, 225 (1978) and references therein. See also G. A. Kachurin, V. A. Bogatyriov, S. I. Romanov, and L. S. Smirnov, in Ion Implantation in Semiconductors 1976, edited by F. Chernow, J. A. Borders, and D. K. Brice, (Plenum, New York, 1977) p. 445.

2) A. A. Grinberg, R. F. Mekhtiev, S. M. Ryvkin, V. M. Salmanov, and I. D. Yaroshetskii, Soviet Physics - Solid State 9, 1085 (1967).

3) M. Bertolotti, L. Stagni, G. Vitali and L. Muzii, J. Appl. Phys. 42, 5893 (1971)

4) F. W. Darby and U.-C. Paek, IEEE J. Quantum Electronics QE 8, 106 (1972).

5) J. F. Giuliani and C. L. Marquardt, J. Appl. Phys. 43, 4993 (1975).

6) J. A. Van Vechten, R. Tsu, F. W. Saris, and D. Hoonhout, to be published in Physics Letters (1979).

7) R. Tsu, J. E. Baglin, T. Tan, and R. J. von Gutfeld, Electrochemical Society Extended Abstracts 79-2, (1979).

8) K. Murakami, M. Kawabe, K. Gamo, S. Namba, and Y. Aoyagi, Phys. Lett. 70A, 332 (1979).

9) R. T. Hodgson, R. Tsu, J. A. Van Vechten, and E. J. Yoffa, Bul. Am. Phys. Soc. 24, 315 (1979).

10) D. H. Auston, C. M. Surko, T. N. C. Venkatesan, R. E. Slusher, and J. A. Goloychenko, Appl. Phys. Lett. 33, 437 (1978); D. H. Auston, J. A. Golovchenko, A. L. Simons, C. M. Surko, and T. N. C. Venkatesan, Appl. Phys. Lett. 34, 777 (1979).

11) D. Hoonhout and F. W. Saris, Physics Letters (in press).

12) J. A. Van Vechten, R. Tsu, and F. W. Saris, Phys. Lett. to be published (1979).

13) J. A. Van Vechten, E. J. Yoffa, R. A. Ghez and R. Tsu, 37th Annual Device Research Conference, Boulder, Colorado, 26 June 1979 Abstract TA1; and to be published.

14) E. J. Yoffa and J. A. Van Vechten, Bull. Amer. Phys. Soc. 24, 315 (1979) and to be published.

15) R. Tsu and S. S. Jha, to be published.

16) C. M. Surko, A. L. Simons, D. H. Auston, J. A. Golovchenko, and R. E. Slusher, Appl. Phys. Lett. 34, 635 (1979).

17) R. Tsu, J. E. Baglin, T. Y. Tan, M. Y. Tsai, K. C. Park and R. T. Hodgson, in Laser-Solids Interactions and Laser Processing - 1978 (AIP Conf. Proc. 50) edited by S. D. Ferris, H. J. Leamy, and J. M. Poate, p. 344.

18) H. R. Shanks, P. D. Maycock, P. H. Sidles and C. G. Danielson, Phys. Rev. 130, 1743 (1963).

19) C. J. Glassbrenner and G. A. Slack, Phys. Rev. ,134A, 1058 (1964).

20) E. J. Yoffa, Laser and Electron Beam Processing of Materials Symposium, Boston, Mass., 27 November 1979, abstract no. I-4; and to be published.

21) J. C. Wang, R. F. Wood, and P. P. Pronko, Appl. Phys. Lett. 33, 455 (1978).

22) P. Baeri, S. U. Campisano, G. Foti and E. Rimini, J. Appl. Phys. 50, 788 (1979).

23) J. C. Schultz and R. J. Collins, Appl. Phys. Lett. 34, 363 (1979).

24) V. Heine and J. A. Van Vechten, Phys. Rev. B 13, 1622 (1976).

25) B. Monemar, R. M. Potemski, M. B. Small, J. A. Van Vechten, and G. R. Woolhouse, Phys. Rev. Lett. 41, 260 (1978).

26) J. R. Troxell, A. P. Chatterjee, G. D. Watkins, and L. C. Kimerling, Phys. Rev. B 19, 5336 (1979) and references therein.

27) J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B 14, 3551 (1976) and references therein.

28) J. A. Van Vechten, Inst. Phys. Conf. Ser. No. 31, 441 (1977).

29) G. D. Watkins, Inst. Phys Conf. Ser. No. 31, 1 (1975).

30) R. A. Swalin, J. Phys. Chem. Solids 18, 290 (1961).

31) R. Andrew and M. Lovato, J. Appl. Phys. 50, 1142 (1979).

32) J. C. Knights, AIP Conf. Proc. 31, 296 (1976).

Appendix E

On the Variation of Semiconductor Band Gaps with Lattice Temperature and with Carrier Temperature When These Are Not Equal*

by

J. A. Van Vechten and M. Wautelet**

I. B. M. Thomas J. Watson Research Center

Yorktown Heights, New York 10598

ABSTRACT

Under conditions of intense optical pumping or electrical injection it is possible to establish a temperature of excited carriers, T_e, larger than the temperature of the lattice, T_L, for periods of time sufficient for many effects to be observed. It is well known that semiconductor band gaps are a function of temperature, but the variation with the two temperatures, T_e and T_L, when these are different seems not to have been discussed previously. Simple thermodynamic arguments may be applied when it is recognized that a band gap is a chemical potential. The simple formula, $\Delta E_{cv}(T_e, T_L) = \Delta H_{cv}(T_L) - T_e \Delta S_{cv}(T_L)$, is deduced. Physically this formula states that the vibronic degeneracy of the electronic states (valence and conduction band or bonding and anti bonding) among which the carriers are distributed with characteristic temperature T_e is determined by the lattice temperature, T_L. Thus when $T_e >> T_L$, anomalously large variations in the gap will occur. It is found that under certain conditions loss of energy from the carrier system to the lattice will cause the density of excited carriers to increase, rather than decrease.

* Supported in part by the Air Force Office of Scientific Research under contract No. F49620-79-C-0077.

** Permanent Address: Faculte des Sciences, Universite de l'Etat, B-7000 Mons, Belgium.

On the Variation of Semiconductor Band Gaps with Lattice Temperature and with Carrier Temperature When These Are Not Equal*

by

J. A. Van Vechten and M. Wautelet**

I. B. M. Thomas J. Watson Research Center

Yorktown Heights, New York 10598

I Introduction

The forbidden energy gap, ΔE_{cv}, of a semiconductor is identically the chemical potential for the formation of unbound electron-hole pairs[1,2]. Indeed, the normal formula for the thermal equilibrium concentration (when non-degenerate),

$$[e_c][h_v] = N_c(T)\, N_v(T) \exp(-\Delta E_{cv}(T)/kT), \tag{1}$$

is an example of the Law of Mass Action[2,3] for the reaction

$$0 \Longleftrightarrow e_c + h_v . \tag{2}$$

Then ΔE_{cv} is, by definition, its standard chemical potential. (Here

$$N_{c,v}(T) = 2\,(2\pi m^*_{c,v} kT/h^2)^{3/2} \tag{3}$$

are the electronic degeneracies, i.e., the effective number of points in the Brillouin zone at which carriers contribute to conduction, for the conduction and valence bands. e_c and h_v represent an electron and a hole excited to the conduction band and valence band edge distribution of states respectively, and m^*_c and m^*_v are their density of states effective masses.)

Consequently, the band gap is equal both to the increase in internal energy upon increase of the carrier density by one electron and one hole under conditions of constant entropy and volume and to the increase in free energy upon increase of the carrier density by one electron and one hole under conditions of constant temperature and pressure;

$$\Delta E_{cv} \equiv \mu_{e,h} \equiv \partial U/\partial n_{e,h}|_{S,V}$$

$$\equiv \partial G/\partial n_{e,h}|_{T,P} \quad . \tag{4}$$

(In a forthcoming book chapter[4] there is given a quantum mechanical derivation of the thermodynamic identity that the increment in free energy measured in thermal experiments is equal to the "no phonon line" value of the internal energy increment measured in optical experiments under normal conditions.)

It has previously been noted[5-7] that there are great advantages when discussing the temperature dependence of band gaps to treat the change in free energy rather than the change in internal energy.

$$\Delta E_{cv} = \Delta G_{cv} = \Delta H_{cv} - T\Delta S_{cv} \quad . \tag{5}$$

The temperature dependence of a free energy is, of course, an entropy, in this case the entropy of the band gap, ΔS_{cv}, or the standard entropy of excitation of an electron across ΔE_{cv}. ΔH_{cv} is the enthalpy of the gap, which must, incidentally, increase with temperature if ΔE_{cv} decreases[2]. The variation of these three thermodynamic variables from T = 0 to the melting point, 1685 K, for Si is shown in Fig. 1, which is borrowed from Ref. 2.

To calculate the temperature dependence of a particular band gap from first principles, one may consider the effect of the excitation of the carriers upon the frequency of the phonon modes rather than the effect of the phonons upon the electronic states[5]. The number of phonon modes excited is always several orders of magnitude greater than the number of carriers excited across the gap. At the melting point of a covalent semiconductor, typically 2.6 times the Debye temperature, all phonon modes are multiply excited whereas only 10^{-4} of the carriers are excited. Whereas simple, tractable and analytic bond charge model[8] formulae are available for the frequency of the various phonon branches in terms of the magnitude of the bond charge and the wave vector, a proper estimate of the effect of the phonons upon the electronic states would have to treat three independent effects - reduction of direct interband scattering (Debye-Waller or Brooks-Yu effect), increase of indirect interband scattering, and

increase of indirect intraband scattering (Fan effect) with increasing temperature - for every one of the phonons in all branches[5]. (The (anharmonic) effect of the variation of lattice parameter with temperature should also be treated, but it is smaller than these effects in the common semiconductors.) It is easier to treat the effect of the excited carriers on the phonon modes.

Under normal conditions the temperature of the phonon distribution, the lattice temperature, T_L, the temperatures which characterize the distribution of carriers within the valence and conduction bands, $T_{e,v}$ and $T_{e,c}$ respectively, and the temperature characterizing the excitation of carriers across the gap, the electron temperature, T_e, are all the same. For this situation ΔE_{cv} can be measured by any of several independent methods from T = 0 K to the melting point, and from these measurements $\Delta H_{cv}(T)$ and $\Delta S_{cv}(T)$ can be deduced. Thurmond has reviewed this data and tabulated the parameters for several semiconductors in Ref. 2.

However, many semiconductor phenomena are observed under "hot electron" conditions for which, due to optical pumping or electrical injection, the three temperatures describing the excited carriers are much greater than that of the lattice[9-12]. As the various direct, indirect and fundamental band gaps, as well as the ionization energies of defects, impurities etc., are all functions of temperature, the condition

$$T_e >> T_L \tag{6}$$

would seem to beg the question of the variation of these chemical potentials with T_e and with T_L: what is $\Delta E_{cv}(T_e, T_L)$? Is ΔE_{cv} a function of $T_{e,v}$ and of $T_{e,c}$?

One approach to the description of nonequilibrium carrier distributions that is well established in the literature is that of the introduction[13] of "quasi-fermi levels". Now, the real fermi level is the chemical potential of electrons and of holes. (Of course, a hole is nothing but the absence of an electron; as chemical potentials are defined by derivatives of total system energies with respect to the number of particles of a given species, electrons and holes

must always have the same chemical potential, or fermi level.) In the (normal, one fitted parameter) "quasi-fermi level" approach, one seeks to describe a nonequilibrium, i.e. $T_e \neq T_L$, concentration of electrons and/or holes by introducing separate values for the electron and hole fermi levels that are adjusted so that the conventional formula gives the actual concentration when the ambient lattice temperature, T_L, and the corresponding value of $\Delta E_{cv}(T_L)$ are inserted. There is no evident thermodynamic significance to these "quasi-fermi levels". The approach is normally used to describe the action of injection lasers or transistors under moderate conditions of injection or optical pumping for which $T_L < T_{e,v} < T_{e,c} << T_e$, because the excited carrier density is not so high that recombination is as fast as energy transfer between carriers or to the lattice. It gives no accurate description of the distribution of carriers within the bands.[13] It could be improved by introducing $T_{e,c}$ or $T_{e,v}$ as a second fitted parameter in addition to the "quasi fermi level", in which case the corresponding band gap and exponential factors ought logically to be employed, so a different value of the "qausi fermi level" would have to be used. It seems quite artificial, but it may be useful when one does not care to develop a thermodynamic description. When none of the four temperatures in the problem are equal, or when, due e.g. to strong electric fields, the distribution of carriers is not accurately described by a thermal distribution with any choice of temperature and chemical potential, this would seem prudent.

However, a proper thermodynamic description is tractable and desirable when the excited carrier concentration is sufficiently high that for times relevant to the experiments of interest, the carrier-carrier interaction is so much stronger than the carrier-phonon interaction that the carrier and phonon systems can be considered to be weakly coupled. Then the carriers will thermalize among themselves before they thermalize with the lattice. Then, when $T_{e,v} \approx T_{e,c} \approx T_e$, we could have $T_e >> T_L$ for times long enough for many interesting effects to be observed. Indeed, Lo and Compaan have shown[14] by direct Raman measure of the surface lattice temperature that T_L does not rise more than about 300 C for crystalline Si exposed to laser pulses well above the threshold energy density for annealing ion implantation

damage.[15-22] They also observed a background scattering characteristic of Raman scattering from carriers with a $T_e >> 2000$ K. Yoffa has shown[23] that, for the excited carrier concentrations that seem to persist during pulsed laser annealing experiments, the carriers would thermalize among themselves much more rapidly than with the lattice, so that the condition discussed here, $T_{e,c} = T_{e,h} = T_e >> T_L$, would obtain. It may also occur for intense electrical injection.

II Derivation

For this case that the carriers may be considered strongly coupled to each other, (so that $T_{e,v} \approx T_{e,c}$) and sufficiently dense that the rates of Auger recombination and impact ionization are rapid on the scale of the experiments of interest (so that $T_{e,c} \approx T_e$), but weakly coupled to the phonons (so that $T_e >> T_L$) for times of interest, a simple formula in terms of the values of $\Delta H_{cv}(T)$ and $\Delta S_{cv}(T)$ measured under the normal condition that $T_e = T_L$ can be deduced from simple consideration of the Law of Mass Action and the number of carriers excited. This formula is

$$\Delta E_{cv}(T_e, T_L) = \Delta H_{cv}(T_L) - T_e \Delta S_{cv}(T_L). \tag{7}$$

To see that this is the correct formula we invite the reader to consider Fig. 2. For simplicity the band structure of the semiconductor has been replaced by a single electronic transition between two electronic states. When the effect of the lattice phonons are added, these two levels become two parabolas (in the harmonic approximation) of vibronic levels. The lattice temperature, T_L, determines the distribution of carriers within whichever vibronic parabola they may be found. At $T_L = 0$ K all carriers must be in the lowest vibronic state because there are no phonons excited. Then the vibronic degeneracy of both states would be the same and the carriers would be distributed between these two parabolas according to the normal Fermi-Dirac expression as a function of the carrier temperature T_e. When $T_L \neq 0$, phonons are excited and carriers in either electronic level are distributed among the corre-

sponding vibronic levels according to the normal Bose-Einstein expression as a function of T_L. In general the lattice stiffness and the phonon energies will be different for the two electronic states, i.e., there is an electron-phonon interaction. Exciting electrons from the valence to the conduction band usually softens the lattice modes of a tetrahedral semiconductor, like Si or GaAs. Cases, like HgTe, where the excitation stiffens the lattice are also known; in such cases ΔE_{cv} should and does have the opposite temperature dependence. These cases are easily understood in terms of the band structures of the various materials.[5]

Fig. 2 is drawn for the usual case that the excitation softens the lattice so the upper parabola is broader than the lower. Consequently the vibronic degeneracy, i.e., the number of states of the total system available to a carrier, is greater for the excited state than for the ground state. By the Law of the Equipartition of States, this will increase the probability of finding the carrier in the excited state for any positive value of T_e.

There are other consequences also. Because the vibronic levels are more closely spaced in the excited electronic state, the excited vibronic levels will begin to be populated to a significant extent at lower values of T_L in the excited state. This means that the average energy of the distribution of carriers in the excited electronic state increases relative to the average distribution in the ground state when ΔE_{cv} decreases with rising temperature, T_L. This difference between the mean energy of the two distributions is just the difference in internal energy for excitation across the gap, which is effectively equal to ΔH_{cv} because pressure times volume change terms are quite negligible in all ordinary circumstances. Indeed, 1 atmosphere is only 0.632×10^{-6} eV/Å^3. (It was noted above that thermodynamics requires ΔH to increase when ΔG decreases.[2])

The standard entropy, ΔS_{cv}, of the excitation reaction is defined in terms of the ratio of the probable number of vibronic states in the two electronic states. As the number of electronic states does not change, this is a function only of the probable number of phonons excited. Therefore, the value of ΔS_{cv}, and thus of ΔH_{cv}, required for $\Delta E_{cv}(T_e, T_L)$ must be

functions of T_L. They will be the same functions of T_L as those measured under normal furnace conditions provided that the number of excited carriers in excess of that which would obtain if the two temperatures were equal is not so large that the effect on the phonons becomes nonlinear in the number of carriers. (The intrinsic carrier concentration at the melting point of Si, and most other common semiconductors, is about 2 X 10^{19} cm^{-3}.) When nonlinearity sets in one would expect the effect to be greater for the excited electronic states because atoms whose bonding electrons have been excited to nonbonding or antibonding electronic states will make larger excursions in their thermal motion than fully bonded atoms and thus will sense the more anharmonic portions of the lattice potential. It follows that our assumption of $\Delta S_{cv}(T_L)$ will underestimate the correct value. This can be seen from Fig. 1 and the fact that ΔS_{cv} does not saturate above the Debye temperature, Θ, as it would in the harmonic approximation[5], but continues to increase moderately all the way to the melting point. At this point it should be clear that the values of the enthalpy and entropy of the gap in Eq. (7) are functions only of the excitation of phonons.

(We might note that, when viewed as a free energy, ΔE_{cv} decreases without limit as T_e increases because the entropy factor in the definition of a free energy, Eq. (7), is multiplied by the temperature relevant to the species being excited, here T_e. When viewed as an enthalpy or internal energy difference, as in optical excitation experiments, ΔE_{cv} decreases because the mean number of phonons excited increases and the optical band gap, when defined as the entropy conserving no-phonon line, connects states with the same number of phonons excited, but each phonon excitation raises the level of the ground electronic state (valence band) more than the excited electronic state (conduction band). The eigenvalues of individual ground and excited vibronic states do not in general approach each other as temperature varies. In the harmonic approximation the eigenvalues do not vary at all.)

The carrier temperature T_e is defined by the distribution of the carriers between the two vibronic systems, the two parabolas, for which the degeneracy is different. If one considers

now just that distribution, as if there were only one temperature in the problem but levels of degeneracy determined in the same way, then it is evident the temperature which multiplies the entropy factor for the free energy, Eq. (7), must be exactly T_e. Thus Eq. (1) becomes,

$$[e_c][h_v] = N_c(T_{e,c})N_v(T_{e,v}) \exp\left(\frac{\Delta S_{cv}(T_L)}{k}\right) \exp\left(\frac{-\Delta H_{cv}(T_L)}{kT_e}\right) \tag{8}$$

(Note that we have here a product of electronic degeneracy factors, N_c and N_v, which are determined by the intraband carrier temperatures, $T_{e,c}$ and $T_{e,v}$, and the vibronic degeneracy factor, $\exp(\Delta S_{cv}/k)$, which is determined by the lattice temperature T_L.) Again, at the very high carrier densities, such as those achieved in laser annealing experiments,[14,18] Yoffa has shown[23] that the three carrier temperatures should approach each other much more rapidly than any of them approaches T_L; the condition $T_{e,v} \approx T_{e,c} \approx T_e >> T_L$ obtains for some time. Eq. (8) simply states that the vibronic degeneracy of the ground and excited electronic states (i.e., the valence and conduction bands or the bonding and antibonding states) is a function of T_L, but the carriers are distributed between these electronic states according to T_e. The ratio of the density of vibronic states available, $\exp(\Delta S_{cv}/k)$, is a factor in the determination of the number of carriers excited. The ratio T_L/T_e does not appear.

III Discussion

One should note that, when $T_e \neq T_L$, the optical no phonon line ("optical gap"), $\partial U/\partial n_{e,h}|_{S,V}$, is not equal to the thermal, or chemical potential, band gap, $\partial G/\partial n_{e,h}|_{T,P}$. This may be seen in Fig. 2 from the fact that the optical no phonon line will depend only on the excitation number of the phonons, and thus T_L; the optical no phonon line will not decrease without limit as T_e increases. However, as the chemical potential and the density of carriers determine diffusion and most other interesting properties, it is the thermal or chemical potential band gap, described by Eq. (7), that is relevant to our discussion.

Another interesting point is that, since $\Delta S_{cv} \rightarrow 0$ as $T_L \rightarrow 0$, as do all other entropies, no major variation of the band gaps should be expected in low energy pulsed laser experiments done with liquid He ambients. Picosecond pulse experiments have commonly been done with samples cooled to liquid He temperatures. If such experiments were repeated with T_L a substantial fraction of Θ, a significant variation of band gaps, optical reflectivity and related phenomena should be observed.

For the same reason, the threshold laser intensity for pusle laser annealing should be sensitive to the initial lattice temperature. If one started from $T_L = 0$ K and the laser pulsed raised it to 300 K, rather than from 300 K to about 600 K, as measured[14] by Compaan and Lo, then from the relative values of ΔS_{cv} (Fig. 1), we see we would need about twice as large a value for T_e to achieve the same band gap and carrier concentration.

When large densities of carriers are excited, the effects of the Exclusion Principle for electrons and holes, i.e. of carrier degeneracy, must be considered. Eqs. (2) and (8) must be replaced with more complicated expressions involving the product of two Fermi-Dirac functions[24],

$$[e_c][h_v] = 4[\frac{kT_e}{2\pi\hbar^2}]^3 (m_c^* m_v^*)^{3/2} \frac{4}{\pi} F_{1/2}(\eta' - \eta_0) F_{1/2}(-\eta'), \tag{9}$$

where, of course,

$$F_{1/2}(\eta) = \int_{x=0}^{\infty} dx \frac{x^{1/2}}{\exp(x-\eta) + 1} \tag{10}$$

Here we have $\eta_0 = \Delta E_{cv}(T_e, T_L)/kT_e$ and $\eta' = E_F/kT_e$, where E_F, is, of course, the fermi level meausred from the valence band edge. When the arguments of the Fermi-Dirac functions are large and negative, these expressions reduce to the Boltzmann expressions given above. When the arguments of these functions become positive (i.e., the band gap becomes negative), the effects of degeneracy become important; for a given concentration of excited

carriers, T_e will actually be larger than the value obtained from Eq. (8). However, such complication does not affect qualitatively the foregoing discussion, which would hold for a system of discrete electronic levels as well as for bands.

One expedient would be to use the Ehrenberg approximation[24,25]

$$F_{1/2}(x) = 2\pi^{1/2} \exp(x)/(4 + \exp(x)), \tag{11}$$

which is accurate when $x < 2.5$, or the approximation[26]

$$F_{1/2}(x) = (4x^{3/2}/3\pi^{1/2}) + \pi^{3/2}/6x^{1/2}, \tag{12}$$

which is accurate when $x > 1.5$. The effect of degeneracy could be accounted for in Eqs. (1) and (8) by simply dividing the classical result by the factor

$$y = (\pi/4) \exp(-\eta_0)/F_{1/2}(\eta'-\eta_0)F_{1/2}(-\eta'), \tag{13}$$

which is plotted in Fig. 3. As long as $-\Delta E_{cv}/kT_e < 2.5$, we may use the Ehrenberg approximation and have

$$y = (16 + 4 \exp(\eta'-\eta_0) + 4 \exp(-\eta') + \exp(-\eta_0))/16\exp(-\eta_0) . \tag{14}$$

Reference to Fig. 1 shows that indeed $x < 2.5$ as long as $T_L < 0.7\ \Theta$ and/or $kT_e < 1$ eV. Therefore, the Ehrenberg approximation should be reasonably adequate for almost all experimental circumstances.

Under pulsed laser annealing conditions, one may contemplate[14,18] the situation where $T_e \approx 10{,}000$ K while $T_L \approx 500$ K. Referring again to Fig. 1, we see that such values of T_e and T_L would produce a decrease in ΔE_{cv} of more than 3 eV from its normal room temperature value of 1.1 eV, so that we would have $x = 1.9$ in the Fermi-Dirac function. (Note that negative values of ΔE_{cv} have significance only in the way they affect the magnitude of the exponential in Eqs. (1) and (8). They do not imply any crossing of eigenvalues.) Thus, pulsed laser annealing conditions can produce anomalously large variations of the thermodynamic band gap. Due to the enormous gradient in the plasma density resulting from the very short absorption lengths of the intense laser pulse, the gradient of the band gap may be very large indeed, of order 10^5 eV/cm (3 eV over 300 nm). As the band gap is smallest where the

carriers are the densest and hottest, this field opposes the expansion of the plasma and may even produce negative carrier diffusion, i.e., produce plasma self-confinement. This point will be discussed further in a separate publication[27].

Another interesting point to consider is the variation of the carrier concentration with time as the carriers do scatter emitting phonons and lose energy to the lattice. Of course, this causes a decrease in T_e and an increase in T_L. The former tends to make $[e_c][h_v]$ decrease through the density of states terms and the denominator of the exponential in Eq. (8). The latter tends to make this product of excited carrier concentrations increase through the $\Delta S_{cv}(T_L/k)$ term in the argument of the exponential. Returning to Fig. 1, we see that, particularly at temperatures below $\Theta/2$, ΔS_{cv} varies rapidly with T_L. If one considers progressively lower values of T_L, the lattice specific heat becomes progressively less. Then a given quantity of energy from the carrier system may produce a progressively larger increase in T_L with a relatively modest decrease in T_e. Thus, we may reach the amusing conclusion that, under certain conditions, a loss of energy from the carrier system will cause the concentration of excited carriers to increase, rather than decrease.

Let us investigating this point more quantitatively. For simplicity let us first treat the non-degenerate expression for $n^2 = [e_c][h_v]$. We obtain (using $T_{e,c} = T_{e,v} = T_e$))

$$\frac{\partial n^2}{\partial T_L} = \frac{n^2}{k}\frac{\partial \Delta H_{cv}}{\partial T_L}\left[\frac{1}{T_L} - \frac{1}{T_e}\right], \tag{15}$$

and

$$\frac{\partial n^2}{\partial T_e} = \frac{n^2}{T_e}\left[3 + \frac{\Delta H_{cv}}{kT_e}\right]. \tag{16}$$

Therefore,

$$\Delta(n^2) = \frac{\partial n^2}{\partial T_L}\Delta T_L + \frac{\partial n^2}{\partial T_e}\Delta T_e = \frac{n^2}{T_e}\times\left[\frac{\partial \Delta H_{cv}}{k\partial T_L}\left(\frac{T_e}{T_L} - 1\right)\Delta T_L + \left[3 + \frac{\Delta H_{cv}}{kT_e}\right]\Delta T_e\right] \tag{17}$$

Using Varshni's equation[28] and Thurmond's data[2], we have

$$\partial \Delta H_{cv} / \partial T_L = 2\alpha\beta^2 T_L / (T_L + \beta)^3, \tag{18}$$

with $\alpha = 4.73 \times 10^{-4}$ eV/°K = 5.49 k and $\beta = 636$ K for Si. Moreover,

$$\Delta T_e = -\Delta T_L \times \frac{C_L}{C_e}, \tag{19}$$

where C_L and C_e are the lattice and carrier specific heats. Around $T_L = 100$ C and at constant pressure, $C_L = 0.18$ cal/gm = 2.55 R for Si.[29]. For a free electron gas at constant volume,

$$C_e = 1.5\ R\ (\pi^2 k T_e / 3\mu_e), \tag{20}$$

where

$$\mu_e = \hbar^2 / 2m\ (3\pi^2 n)^{2/3}. \tag{21}$$

Let us make the most conservative estimate for Δn^2 by choosing the maximum plausible value of μ_e, which would obtain if all the 2×10^{23}/cm^3 valence electrons would be participating in the gas, so $\mu_e = 12.5$ eV and we find $C_e|_V = 0.39$ R. Then we would have $C_e|_P = 1.39$ R, the carrier specific heat at constant pressure, to be compare with the lattice specific heat $C_L = 2.55$ R. Then, evaluating Eq. (16) we find,

$$\Delta n^2 = 53\ \Delta T_L\ n^2 / T_e, \tag{22}$$

i.e., while these extreme conditions remain, the loss of energy from the carrier system to the lattice causes the density of excited carriers to increase rather than decrease. Again, this is simply the result of increasing the vibronic degeneracy of the excited electronic states, which increases the probability that they be occupied.

Of course, degeneracy will generally have significant effect under the conditions relevant to this discussion. Physically, it is clear that an effect of degeneracy is to increase the magnitude of the phenomenon just discussed, whereby loss of energy from the carriers to the lattice causes the concentration of excited carriers to increase. This is simply because degeneracy decreases the rate of change of n^2 with T_e, and thus lessens the effect of decreas-

ing T_e. However, the effect of increasing T_L with the transferred energy is to increase the number of vibronic states available, and thus serves to releave the restriction of the Exclusion Principle.

Let us denote the correct value of $[e_c][h_v]$, taking degeneracy into account, as $n^2 y$, where y is defined in Eq. (13). If we then use the Ehrenberg approximation, Eq. (11), for the Fermi-Dirac function in order to take account of electronic degeneracy, then we find Eq. (16) should be modified to,

$\Delta(n^2/y) =$

$$\frac{n^2}{yk}\frac{\partial \Delta H_{cv}}{\partial T_L}\left[\frac{1}{T_L}-\frac{1}{T_e}\right](1 + f(\eta_0,\eta')/y)\Delta T_L + \frac{n^2}{yT_e}\left[3 + (1 + f(\eta_0,\eta')/y)\frac{\Delta H_{cv}}{kT_e}\right]\Delta T_e, \quad (23)$$

and

$$F(\eta_0,\eta') = \left[4 + \frac{\eta_0-\eta'}{\eta_0}\exp(-\eta') + \frac{\eta'}{\eta_0}\exp(-(\eta_0-\eta'))\right]/4. \quad (24)$$

$F(\eta_0,\eta')$ is always positive and becomes large when η_0 becomes large and negative, i.e., when the band gap becomes large and negative, as will occur under intense excitation, as in pulsed laser annealing. If the approximation of Eq. (12) were used, a much more complicated expression would be obtained, but the qualitative effect would be the same. The whole of the ΔT_L term, which causes an increase in carrier concentration would be multiplied by a factor greater than unitiy, while only a portition of the ΔT_e term, which causes a decrease in carrier concentration, is multiplied by the same factor. Note that the portion of that term which is multiplied by the factor is the smaller of the two under pulsed laser annealing conditions.

References

1) H. Brooks, Advan. Electric Electron. Phys. 7, 85 (1955).

2) C. D. Thurmond, J. Electrochem. Soc. 122, 1133 (1975).

3) Cf., e.g., C. Kittel, Thermal Physics (John Wiley and Sons, New York, 1969) pp. 348.

4) J. A. Van Vechten, Handbook of Semiconductors (North Holland, Amsterdam, 1980) edited by S. P. Keller, pp. 1.

5) V. Heine and J. A. Van Vechten, Phys. Rev. B 13, 1622 (1976).

6) J. A. Van Vechten, Phys. Rev. B 13, 946 (1976).

7) J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B 14, 3539 (1976).

8) R. M. Martin, Phys. Rev. 186, 871 (1969).

9) K. Betzler, Solid State Comm. 15, 1837 (1974).

10) J. Shah, Solid State Electron. 71, 43 (1978).

11) E. O. Gobel and G. Mahler, Festkorper Probleme 19, 105 (1979).

12) R. Ulbrich, Solid State Electr. 21, 51 (1978).

13) Cf., e.g., R. A. Smith, SEMICONDUCTORS, Second Edition (Cambridge University Press, Cambridge, 1978) pp. 176.

14) H. W. Lo and A. Compaan, Phys. Rev. Lett. 44, 1604 (1980).

15) G. H. Schwuttke, J. K. Howard, and R. F. Ross, U. S. Patent 3,585,088 filed 10/18/1968; issued 6/15/1971.

16) I. B. Khaibullin, E. I. Shtyrkov, M.M. Zaripov, R. M. Bayazitov and M. F. Galjautinov, Radiat. Eff. 36, 225 (1978), and references therein.

17) G. A. Kachurin, V. A. Bogatyriov, S. I. Romanov, and L. S. Smirnov, in "Ion Implantation in Semiconductors, 1976", eds. F. Chernow, J. A. Borders and D. K. Brice (Plenum, New York 1977) p. 445, and references therein.

18) J. A. Van Vechten, R. Tsu, F. W. Saris, and D. Hoonhout, Phys. Lett. 74a, 417 (1979), and references therein; J. A. Van Vechten, J. de Phys. 41, C4-15 (1980); R. Tsu and S. S. Jha, J. de Phys. 41, C4-25 (1980).

19) C. W. White, J. Narayan, and R. T. Young, Science 204, 225 (1978).

20) P. Baeri, S. U. Compisano, G. Foti, and E. Rimini, Phys. Rev. Lett. 41, 1246 (1978).

21) D. H. Auston, C. M. Surko, T. N. C. Venkatesan, R. E. Slusher, and J. A. Golovchenko, Appl. Phys. Lett. 33, 437 (1978).

22) K. Murakami, M. Kawabe, K. Gamo, S. Namba, and Y. Aoyagi, Phys. Lett. 20a, 332 (1979).

23) E. J. Yoffa, Phys. Rev. B 21, 2415 (1980).

24) Cf., e.g., R. A. Smith, op. cit. (Ref. 13), pp. 77.

25) W. Ehrenberg, Phil. Trans. A 237, 67 (1938).

26) K. Seeger, "Semiconductor Physics" (Springer-Verlag, New York 1973) p. 42.

27) M. Wautelet and J. A. Van Vechten, to be published (accompanying paper).

28) Y. P. Varshni, Physica 39, 149 (1967).

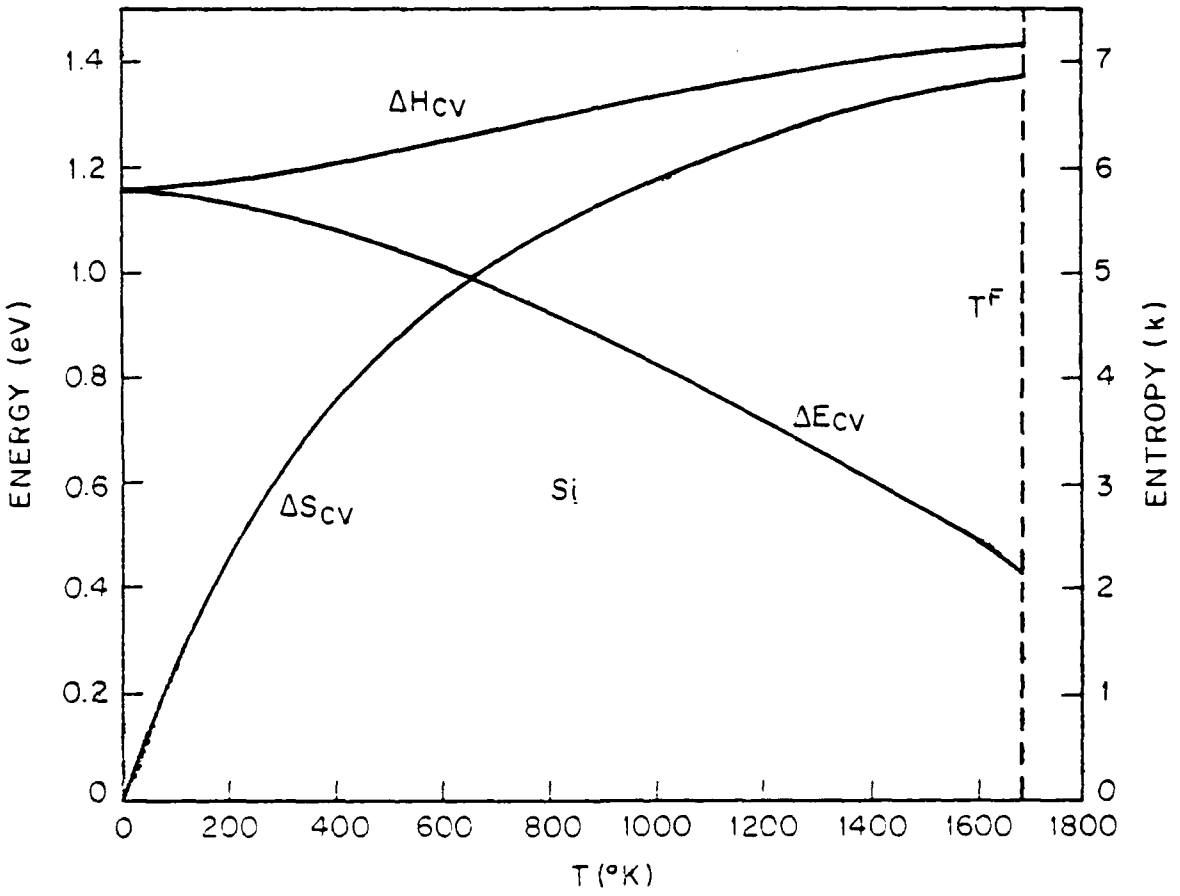
29) Cf., e.g., W. R. Runyan, "Silicon Semiconductor Technology" (McGraw-Hill, New York, 1965) p. 210.

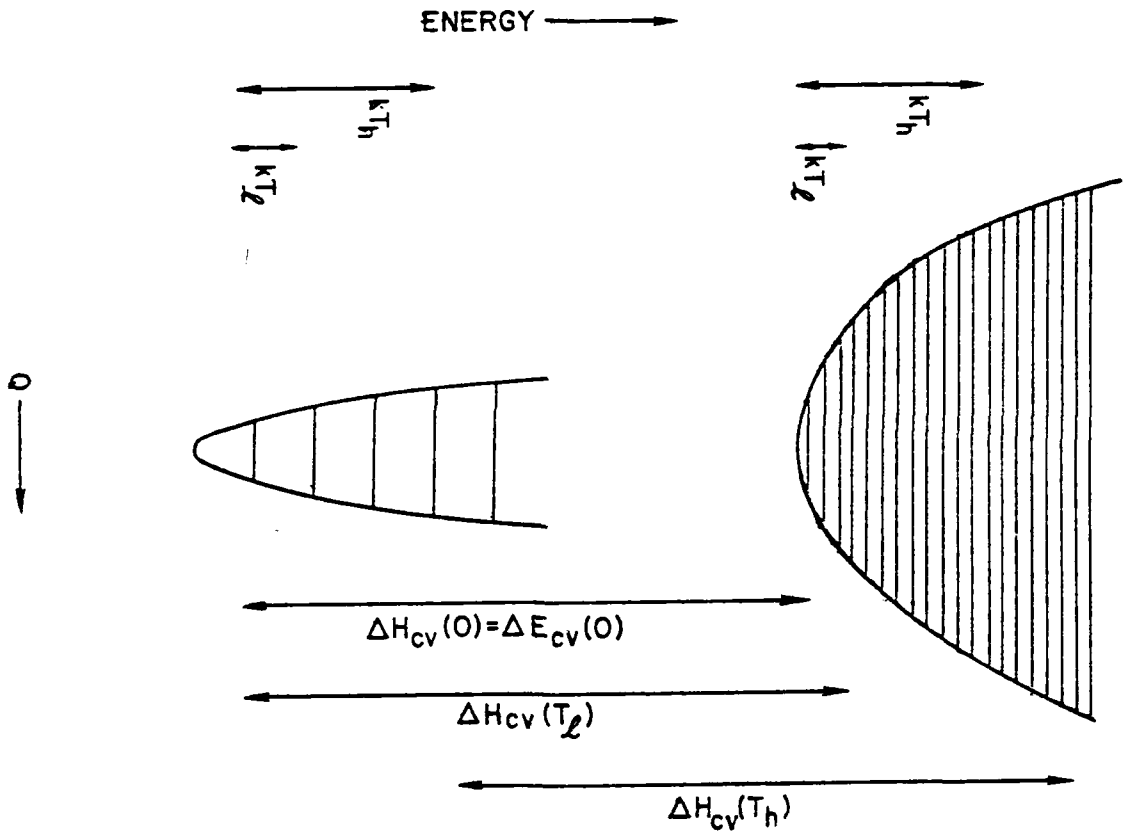
Figure Captions

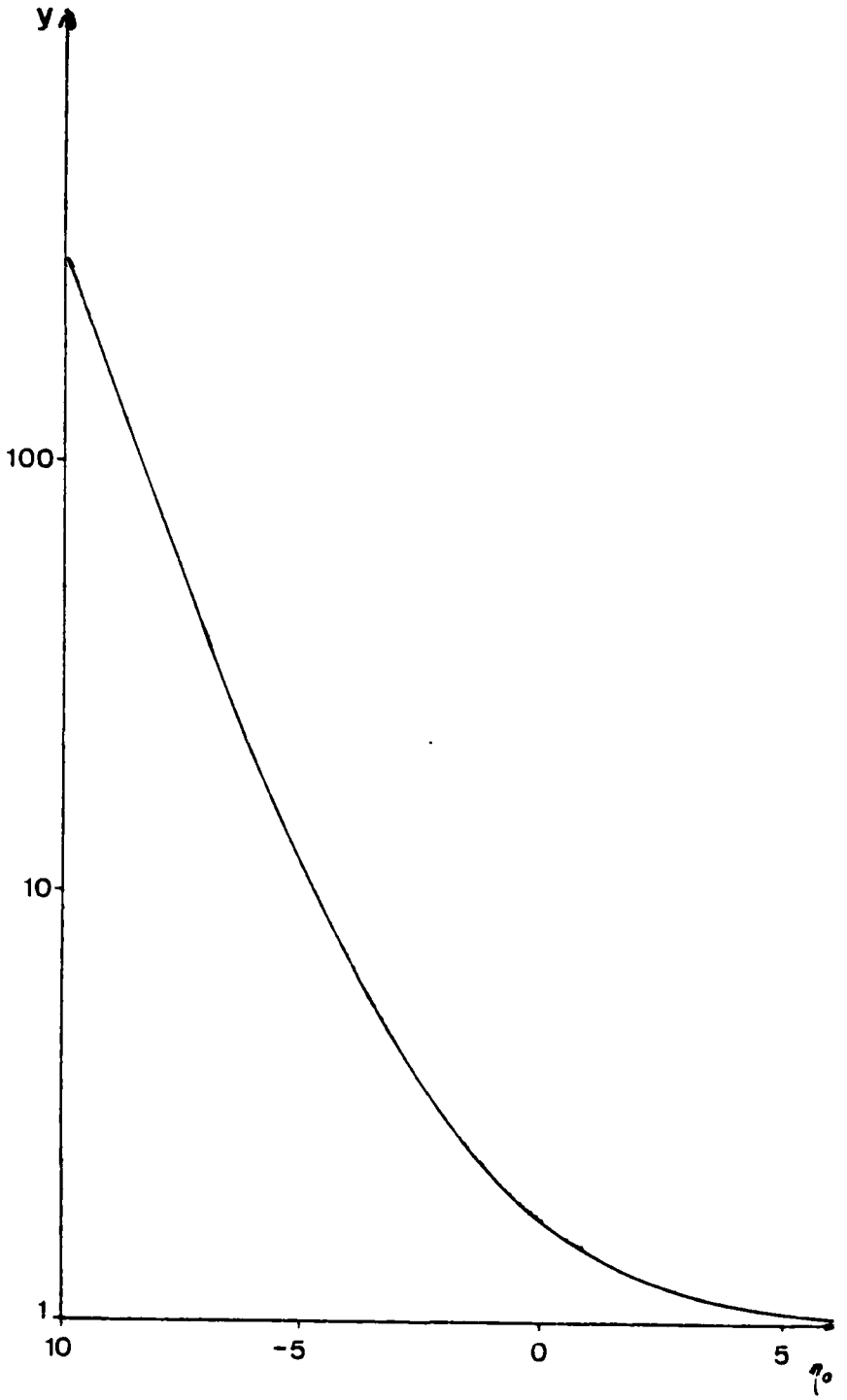
Fig. 1 Variation with temperature of the fundamental band gap, ΔE_{cv} of Si. As ΔE_{cv} is the standard chemical potential for the creation of (unbound) electron-hole pairs, it is equal to a free energy. The corresponding standard enthalpy and standard entropy of this reaction, ΔH_{cv} and ΔS_{cv}, are also shown. This figure is borrowed from Thurmond, Ref. 2.

Fig. 2 Configuration coordinate diagram to illustrate how it is that when the entropy of formation of electron-hole pairs is positive (as it is in Si and most other semiconductors), then, in the harmonic approximation: the vibronic eigenvalues change not at all; the enthalpy difference between the two electronic states must increase with carrier temperature, T_e, to a saturation value; and the free energy difference between the two electronic levels will decrease without limit as the T_e increases. T_l and T_h denote lattice temperatures which are respectively low and high with respect to the Debye temperature. Also, the value of ΔS_{cv}, and therefore of ΔH_{cv}, is entirely a function of the number of vibrational modes excited for each electronic state. These entropies and enthalpies will always be functions of T_L, and will, until the carrier density becomes great enough to significantly alter phonon frequencies, be functions only of T_L.

Fig. 3 Variation of degeneracy correction factor, y, from Eq. (13) with $\eta_0 = \Delta E_{cv}(T_e,T_L)/kT_e$, into the range of very high excitation, as during pulsed laser annealing, where the band gap becomes negative.


ENERGY (eV)
ENTROPY (k)
T (°K)
ΔHcv
ΔEcv
ΔScv
Si
T^F
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
7
6
5
4
3
2
1
0
0
200
400
600
800
1000
1200
1400
1600
1800


ENERGY
kT_h
kT_ℓ
kT_h
kT_ℓ
D
ΔH_cv(0)=ΔE_cv(0)
ΔH_cv(T_ℓ)
ΔH_cv(T_h)


y
100
10
1
10
-5
0
5
η₀

Appendix F

Carrier Diffusion in Semiconductors Subject to Large Gradients of Excited Carrier Density*

by

M. Wautelet** and J. A. Van Vechten

I. B. M. Thomas J. Watson Research Center

Yorktown Heights, New York 10598

ABSTRACT

Under conditions of intense optical pumping or electrical injection it is possible to excite such a dense plasma of electrons and holes that these carriers interact among themselves much more rapidly than they interact with the lattice phonons. In this circumstance there will be established a temperature of excited carriers, T_e, much larger than than the temperature of the lattice, T_L, for periods of time sufficient for many effects to be observed. In the accompaning paper it is shown that the temperature dependence of the forbidden band gap is then described by $\Delta E_{cv}(T_e, T_L) = \Delta H_{cv}(T_L(T_L) - T_e \Delta S_{cv}(T_L)$. With $T_e >> T_L$, this means that anomalously large changes in the the band gap will occur and, if also there are large gradients of T_L present in the sample, there will be very large gradients in the band gap. This means there will be very large gradients in the chemical potentials of the excited electrons and holes. As the band gap tends to be least where the carriers are most dense and T_L the greatest, these chemical potential gradients oppose the normal outward diffusion of the excited carriers. We find that, under certain conditions, this effect may be sufficient to produce "negative diffusivity" or carrier self confinement.

* Suppored in part by project IRIS of the Belgian Ministry of Science and in part by the Air Force Office of Scientific Research under Contract No. F49620-79-C-0077.

** Permanent Address: Faculte des Sciences, University de l'Etat, B-7000 Mons, Belgium.

Carrier Diffusion in Semiconductors Subject to Large Gradients of Excited Carrier Density*

by

M. Wautelet** and J. A. Van Vechten

I. B. M. Thomas J. Watson Research Center

Yorktown Heights, New York 10598

I Introduction

The forbidden band gap, ΔE_{cv}, of a semiconductor is identically the standard chemical potential for the formation of unbound electron-hole pairs.[1,2] (The standard chemical potential is that part of the total chemical potential which does not depend explicitly on the concentration of the species.) It is also well known to be a function of temperature.[2,3] Under "hot electron" conditions of intense optical pumping, irradiation, or electrical injection, the temperature, T_L, which characterizes the lattice phonons, will not be equal to the temperatures which characterize the distribution of carriers excited within the valence and conduction bands, $T_{e,v}$ and $T_{e,c}$, and the temperature which characterizes the excitation of carriers across the gap, T_e. Carrier - carrier interaction is generally so strong that we may assume

$$T_{e,c} = T_{e,v}. \tag{1}$$

When the excited carrier density is sufficiently high that the rates of Auger recombination and impact ionization are rapid on the scale of the experiment of interest, while carrier lattice interaction is slow on that scale, we reach an interesting regime in which the carrier system may be considered to be in quasiequilibrium with itself and weakly coupled to the lattice. Thus,

$$T_e = T_{e,c} = T_{e,v} >> T_L, \tag{2}$$

for times relevant to the experiment under consideration.

It has been argued[4-8] for some time that pulsed laser annealing[4-14] is a nonthermal phenomenon. More recently it has been argued[7,8] that it primarily involves a plasma of excited electrons and holes at a density of order 10^{21}/cm^3 and $T_e >> T_L \approx 300$ C. The argument regarding the thermal or nonthermal nature of pulsed laser annealing was finally settled, when Lo and Compaan showed[14] by direct Raman measure of surface lattice temperature that T_L does not rise more than 300 C for at least 50 ns after the end of a laser pulse of intensity just below the threshold for damage and more than twice the threshold for normal annealing. Therefore, the electron-hole plasma created by the absorption of this light must retain the energy of this pulse and is heated to several 1000 C. Indeed, Lo and Compaan observed a background scattering characteristic of Raman scattering from carriers with $T_e >>$ 2000 K. Yoffa's calculations[13] show that when an excited plasma is this dense, the condition of Eq. (2) should be expected to prevail. What is most remarkable about Lo and Compaan's experimental result[14], and indeed the whole pulsed laser annealing phenomena, is that it prevails for such a long time[7,8].

When this is the case, we have shown in the accompanying paper[15] that the band gap is described by the simple formula,

$$\Delta E_{cv}(T_e, T_L) = \Delta H_{cv}(T_L) - T_e \Delta S_{cv}(T_L), \tag{3}$$

where ΔH_{cv} and ΔS_{cv} are the enthalpy and standard entropy of the excitation of the carrier across the gap[1,2]. These thermodynamic parameters are functions of T_L and should be affected only moderately by T_e until the carrier density becomes so high that nonlinearites become important. They have been measured under normal furnace conditions, $T_e = T_L = T$, from T = 0 K to temperatures near the melting points of several semiconductors. This data has been reviewed and tabulated by Thurmond in Ref. 2. Experimental values for Si to its melting point are shown in Fig. 1, which is borrowed from Ref. 2.

At the very high carrier densities relevant to pulsed laser annealing, the concentration of excited electrons, $[e_c^-] = [h_v^+] \equiv n$, the concentration of excited holes. Therefore, we may

consider only the ambipolar diffusion of the neutral plasma rather than the diffusion of electrons and of holes separately. Gradients of T_e and of T_L will produce a gradient of ΔE_{cv}, which is equivalent to an internal field tending to drive the plasma toward the regions where ΔE_{cv} is minimal. As T_e and T_L are greatest where n is greatest and ΔE_{cv} is least, the effect of the internal field is to oppose the normal outward diffusion of the plasma from the region where it was created.

We will here show that under the extreme conditions of pulsed laser annealing this internal field due to ∇T_L and ∇T_e may be sufficient to completely overcome the outward diffusion and cause it to contract. Although we have not solved the problem of the variation of the plasma density with time, it appears that this confinement phenomenon will persist about as long as ∇T_L remains large. That would be comparable to the time required for an equivalent amount of heat to diffuse away, i.e., for times of order 100 ns.

II Derivation

The effect of gradients of the band gap, whether induced by thermal or by composition gradients, has, of course, been treated previously in the literature.[16-18] It is particularly significant to the operation of double heterostructure lasers, where it causes both electrons and holes to be confined in the active layer.[18]

To simplify the discussion, we shall take the ratio of the laser annealed spot diameter to the depth of penetration of the heat and the carriers to be sufficiently large that the problem may be treated as one dimensional. Thus, we only consider diffusion perpendicular to the surface, which direction we denote z. Following the notation of Stratton[16], let us denote the current of electron-hole pairs at a particular point, z, near the surface of the sample in the z direction as J. Further,

$$J = J_F + J_D, \tag{4}$$

where J_F is the contribution due to the gradient of the band gap, which drives pairs up towards the surface where their density is greatest,

$$J_F = n\mu F, \tag{5}$$

and where J_D is the contribution due to normal diffusion, which drives pairs down away from the surface,

$$J_D = - D\, dn/dz. \tag{6}$$

(We neglect the fact that D may vary with z in this rather qualitative discussion.) Here, F is the field acting on the pairs due to the variation of the band gap,

$$F = -\frac{d\Delta E_{cv}}{dz} = -\frac{d\Delta H_{cv}}{dT_L}\frac{dT_L}{dz} + \Delta S_{cv}\frac{dT_e}{dz} + T_e\frac{d\Delta S_{cv}}{dT_L}\frac{dT_L}{dz}. \tag{7}$$

D is the ambipolar diffusivity of the pairs, and μ is their mobility. D and μ are connected by the Einstein relation,

$$\mu = DF_{-1/2}(\eta'-\eta_0)F_{-1/2}(-\eta')/kT_eF_{1/2}(\eta'-\eta_0)F_{1/2}(-\eta'), \tag{8}$$

where $F_{-1/2}$ and $F_{1/2}$ are the Fermi-Dirac functions and $\eta_0 = -\Delta E_{cv}/kT_e$ and $\eta' = E_F/kT_e$, with Fermi energy, E_F measured from the valence band edge. For the present we approximate Eq. (8) with the non-degenerate formula,

$$\mu = D/kT_e. \tag{9}$$

The error introduced thereby is only a factor of 2.6 even at $\eta_0 = 2.5$.

We use the Varshni equations[19] and the data of Thurmond[2] for $\Delta H_{cv}(T_L)$ and $\Delta S_{cv}(T_L)$. These give

$$d\Delta H_{cv}/dT_L = 2\alpha\beta^2T_L/(T_L + \beta)^3 \tag{10}$$

and

$$d\Delta S_{cv}/dT_L = 2\alpha\beta^2/(T_L + \beta)^3, \tag{11}$$

with $\alpha = 4.73 \times 10^{-4}$ eV/°K = 5.49 k and $\beta = 636$ K for Si.

Thus, Eq. (4) becomes

$$\frac{J}{D} = \frac{2n\alpha\beta^2}{k(T_L+\beta)^3}\left[1-\frac{T_L}{T_e}\right]\frac{dT_L}{dz} + \frac{n\Delta S_{ev}}{kT_e}\frac{dT_e}{dz} - \frac{dn}{dz} \qquad (12)$$

The first term, which vanishes when $T_e = T_L$, comes from combining the contributions from $d\Delta H_{ev}/dT_L$ and $d\Delta S_{ev}/dT_L$. (Positive values of J/D imply the plasma is confined and driven to higher densities.) Note that T_e is a function of n because that is how it is defined.

III Discussion

Initially after a short laser pulse both dn/dz and dT_e/dz would be given by the more-or-less exponential distribution of pairs produced by the absorption of the light, but, if confinement begins, they both will grow larger. The question of a proper determination of dT_L/dz is difficult because of the open question of the variation with plasma density of the rate of transfer of energy from the plasma to the lattice. Here we shall simply assume dT_L/dz is proportional to dn/dz in the period immediately following the laser pulse. As no mechanism for confining the flow of heat, i.e., the phonons, has been suggested, it seems clear that after some time dT_L/dz will grow small.

Let us now evaluate Eq. (12) using parameters consistent with the measurements[14] of Lo and Compaan for the period immediately following the laser pulse. We take $T_L = 573$ K so that $\Delta S_{ev} = 4.0$ k. Then Eq. (12) becomes,

$$J/D = 2.5 \times 10^{-3}\, n\,[1 - T_L/T_e]\, dT_L/dz + 4.0n/T_e\, dT_e/dz - dn/dz. \qquad (13)$$

We further assume initial penetration to have a characteristic depth of 100 nm, so that with $n = 10^{21}$ cm^{-3}, $dn/dz = 10^{26}$ cm^{-4}, $dT_L/dz = 3 \times 10^{7}$ °K/cm, $T_e \approx 10^{4}$ K, and $dT_e/dz = 10^{4}$ °K/cm. With these values, Eq. (12) now becomes,

$$J/D = [\,0.75 + 4.0 - 1\,] \times 10^{26}\ \text{cm}^{-4} = +3.75 \times 10^{26}\ \text{cm}^{-4} \qquad (14)$$

This positive value is large enough that the error due to the neglect of degeneracy in using Eq. (9) instead of (8) could not change the sign of the result. We further note that the dominant positive term is that involving dT_e/dz, which will grow larger with time if confinement occurs, and not that involving dT_L/dz, which will grow smaller after some time. Finally, as was noted in the accompanying paper[15], the effect of including the effect of degeneracy in the relation between T_e and n is to increase T_e relative to n, so that dT_e/dz will also be increased relative to dn/dz.

IV Conclusions

We conclude that there is no doubt that for the extreme conditions achieved in some pulsed laser annealing experiments there will be an initial period in which the laser generated plasma of electron-hole pairs will not diffuse down its concentration gradient but will be driven to higher densities by the band bending induced by the gradient of the band gap. To understand the time development of this confinement, we will have to develop an understanding of the variation with plasma density of the rate of transfer of energy from the carrier system to the lattice. While progress is being made in this direction[8,13,20,21] a full understanding of this problem, and of the long persistence of the laser induced plasma, has not yet been attained.

References

1) H. Brooks, Advan. Electric Electron. Phys. 7, 85 (1955); see also H. M. James in "Photoconductivity Conference Held at Atlantic City 1954", edited by R. G. Breckenridge, B. R. Russell, and E. E. Hahn (John Wiley, New York, 1956) p. 204.

2) C. D. Thurmond, J. Electrochem. Soc. 122, 1133 (1975).

3) V. Heine and J. A. Van Vechten, Phys. Rev. B 13, 1622 (1976).

4) G. H. Schwuttke, J. K. Howard, and R. F. Ross, U. S. Patent 3,585,088 filed 10/18/1968; issued 6/15/1971.

5) I. B. Khaibullin, E. I. Shtyrkov, M.M. Zaripov, R. M. Bayazitov and M. F. Galjautinov, Radiat. Eff. 36, 225 (1978), and references therein.

6) G. A. Kachurin, V. A. Bogatyriov, S. I. Romanov, and L. S. Smirnov, in "Ion Implantation in Semiconductors, 1976", eds. F. Chernow, J. A. Borders and D. K. Brice (Plenum, New York 1977) p. 445, and references therein.

7) J. A. Van Vechten, R. Tsu, F. W. Saris, and D. Hoonhout, Phys. Lett. 74a, 417 (1979), and references therein.

8) J. A. Van Vechten, J. de Phys., 41, C4-15 (1980) (Proceedings of Mons Conference on Laser Induced Nucleation in Solids, Mons Belgium 10/4-6/79).

9) C. W. White, J. Narayan, and R. T. Young, Science 204, 225 (1978).

10) P. Baeri, S. U. Compisano, G. Foti, and E. Rimini, Phys. Rev. Lett. 41, 1246 (1978).

11) D. H. Auston, C. M. Surko, T. N. C. Venkatesan, R. E. Slusher, and J. A. Golovchenko, Appl. Phys. Lett. 33, 437 (1978).

12) K. Murakami, M. Kawabe, K. Gamo, S. Namba, and Y. Aoyagi, Phys. Lett. 20a, 332 (1979).

13) E. J. Yoffa, Phys. Rev. B 21, 2415 (1980).

14) H. W. Lo and A. Compaan, Phys. Rev. Lett. 44, 1604 (1980).

15) J. A. Van Vechten and M. Wautelet, to be published (accompanying paper).

16) R. Stratton, IEEE Trans. Electron. Devices, ED-19, 1288 (1972).

17) P. N. Butcher and L. Friedman, J. Phys. C. 10, 3803 (1977).

18) F. Stern, Phys. Rev. 148, 186 (1966).

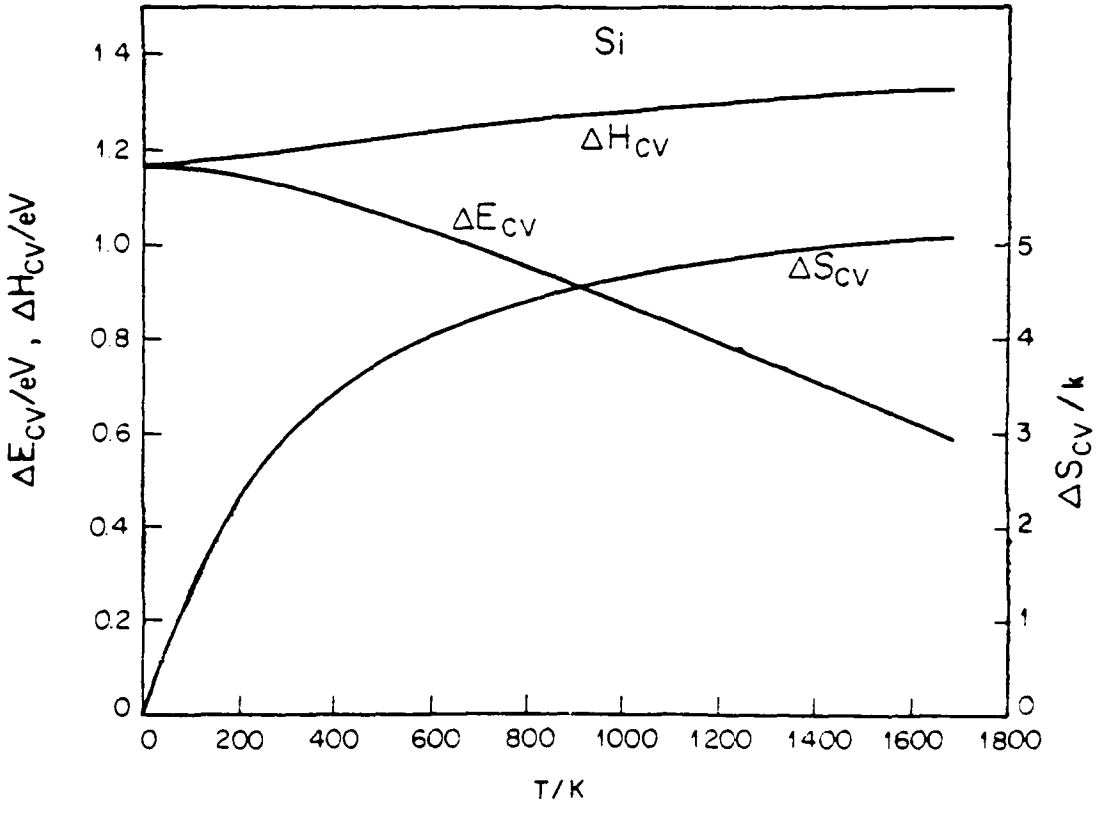
19) Y. P. Varshni, Physica 39, 149 (1967).

20) J. A. Van Vechten, R. Tsu, and F. W. Saris, Phys. Lett. 74a, 422 (1979).

21) R. Tsu and S. S. Jha, J. de Phys. 41, C4-25 (1980) (Proceedings of Mons Conference on Laser Induced Nucleation in Solids, Mons, Belgium 10/4-6/79).

Figure Captions

Fig. 1 Variation with temperature of the fundamental band gap, ΔE_{cv} of Si. As ΔE_{cv} is the standard chemical potential for the creation of (unbound) electron-hole pairs, it is equal to a free energy. The corresponding standard enthalpy and standard entropy of this reaction, ΔH_{cv} and ΔS_{cv}, are also shown. This figure is borrowed from Thurmond, Ref. 2.


Si
ΔHCV
ΔECV
ΔSCV
ΔECV/eV , ΔHCV/eV
ΔSCV / k
T/K
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1
2
3
4
5
0
200
400
600
800
1000
1200
1400
1600
1800

Appendix G

PHYSICAL REVIEW B VOLUME 21, NUMBER 6 15 MARCH 1980

Dynamics of dense laser-induced plasmas

Ellen J. Yoffa
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 17 August 1979)

Calculations have been made to determine the influence of a dense plasma of hot electrons and holes on the primary channels of energy relaxation and redistribution of photoexcited carriers in Si, particularly collisions between carriers, plasmon emission, impact ionization, phonon emission, and carrier diffusion. At high carrier densities, Auger recombination is sufficiently fast to ensure that the electrons and holes rapidly reach quasiequilibrium with a common quasi-Fermi level at a temperature which is lowered by the partitioning of energy into thermally excited plasmons. The appropriate dielectric function has been calculated. At sufficiently high temperatures and carrier densities, energy can diffuse at a rate that is comparable to (and, in some cases, faster than) the rate at which the energy is transferred to the lattice. The steady-state carrier density and temperature, and consequently the ultimate extent to which the lattice is heated, depend critically on the parameters of the exciting laser.

I. INTRODUCTION

It has recently been argued that simple-melting or strictly thermal models for pulsed laser annealing of ion-implanted and amorphous Si cannot provide consistent explanations for a large body of experimental evidence.[1-3] It has been suggested as an alternative possibility that annealing is achieved in the presence of an electron-hole plasma.[1,3] In this paper we shall examine a related topic: the dynamics of a dense laser-induced plasma and, in particular, the influence of high concentrations of hot carriers on the rate at which energy is transferred from the laser to the silicon lattice. Previous calculations[4-8] have been performed which assume that the laser energy is transferred to the lattice in the same region in which it is initially absorbed. We shall demonstrate that under certain conditions carrier densities and temperatures are so high that during the laser pulse energy can diffuse from the irradiated volume faster than it heats the lattice. In addition, we shall find that for sufficiently high carrier densities, the phonon emission rate is itself screened. Most theoretical and experimental investigations of energy relaxation on psec to nsec time scales of hot, photoexcited carriers have dealt with carrier densities in Ge (Refs. 9 and 10) and GaAs (Refs. 11–15) which are lower than those at which one would expect the above-mentioned effect to occur.

In the following sections, we shall examine the rate at which energy is given to the carriers by the laser, the rates at which carrier collisions redistribute this energy, and the rate at which the energy is transferred to the lattice. In order to compare the relative importance of competing energy-transfer mechanisms, we shall refer our calculations to a typical laser annealing experiment[16] with laser wavelength $\lambda_L = 0.53$ μm, incident power density $P = 10^7$ W cm^{-2}, and laser pulse duration $\tau_L = 10$ nsec. These parameters correspond to an incident energy density of 1 J cm^{-2}. Although annealing is often performed in the presence of high implanted donor or acceptor concentrations, we shall assume that any extrinsic carrier densities introduced are small compared with the photoexcited densities.

In Secs. II and III we shall discuss carrier creation and the subsequent thermalization and recombination. The interdependence of the resulting carrier density and energy is examined in Sec. IV. In Sec. V we calculate the effect of a hot, dense plasma on the rate at which energy is transferred to the lattice by phonon emission. In Sec. VI we demonstrate the importance of carrier diffusion in determining the ultimate extent to which the lattice is heated. Finally, in Sec. VII we discuss the implications of our analysis.

II. CARRIER THERMALIZATION

The incident laser energy is absorbed either by electron-hole pair creation or by free-carrier excitation. Initially, with few carriers present, the former process dominates, so that near the silicon surface the photon absorption rate $g = P(1-R)\,\delta/(\hbar\omega_L)$, where P is the incident laser power per unit area, R is the reflectivity of the sample ($R \approx 0.5$), $\hbar\omega_L$ is the photon energy, and δ is the absorption length ($\delta \sim 10^{-5}$ cm at $\hbar\omega_L = 2.3$ eV). Electron-hole pairs are created via indirect absorption processes involving the emission and absorption of phonons. Because phonon energies are much smaller than the photon energy, the amount of energy transferred to the lattice during absorption is negligible in comparison to the total amount absorbed. The rise in

carrier density leads, in turn, to increased free carrier absorption. By the same argument as above, we can neglect the energy transferred to the lattice by this process also. The net result is consequently the production of hot electrons and holes which then thermalize with the rest of the carriers and eventually with the lattice. For the example we are considering, $g \sim 10^{31}$ cm^{-3} sec^{-1}, so that carrier densities exceeding 10^{19} cm^{-3} are achieved in times much less than the laser pulse duration. We shall therefore look at carrier thermalization at concentrations greater than 10^{19} cm^{-3}.

Primary channels for energy relaxation of the hot carriers are collisions with the other carriers, plasmon production, and phonon emission. Another important mechanism is electron-hole pair production by impact ionization. (This process is the inverse of Auger recombination, and will be discussed in detail in Sec. III.) Of these processes, all but phonon emission involve primarily the redistribution of the carrier energy among the electrons and holes with negligible amounts of energy transferred to the lattice.

For an electron having an energy E above the conduction-band minimum, the rate of energy loss due to collisions with the N_e other electrons is given by[17]

$$\left(\frac{\partial E}{\partial t}\right)_{e-e} = -\frac{4\pi N_e e^4 c f}{\epsilon_0^2 (2m_e^* E)^{1/2}}, \tag{1}$$

where e and m_e^* are the electron charge and effective mass, and ϵ_0 is the bulk dielectric constant. The energy lost by the hot electron is then shared by all of the carriers. We shall demonstrate in Sec. V that at very high carrier densities, the detailed dependence of ϵ on wave vector $\vec{q}$ and frequency ω becomes important, modifying the simple form of Eq. (1). For the purposes of comparing the relaxation rates, however, simple expressions containing ϵ_0 are adequate. c is a factor between 1 and 2 which accounts for the influence of spin and exchange on a scattering event.[17] f is of order 1, and depends on the distribution of electrons with which the hot electron interacts. Clearly, the importance of the particular distribution diminishes as the energy of the hot electron increases.[17] An expression similar to Eq. (1) applies to electron-hole collisions. As we shall find in Sec. VII, N_e for our example is $\sim 10^{20}$ cm^{-3}, so that with $E \approx 1$ eV, $\epsilon = 11.8$, and $m_e^* = 0.33m$, $(\partial E/\partial t)_{e-e}$ for our example is roughly -10^{14} eV/sec.

Energy relaxation by plasmon emission is possible when E is greater than $\hbar\omega_p$, where $\hbar\omega_p$ is the plasmon energy,

$$\hbar\omega_p = \hbar\left(\frac{4\pi N_e e^2}{\epsilon_0 m_r^*}\right)^{1/2}, \tag{2}$$

with $m_r^* = 0.15m$ the reduced electron-hole effective mass.[18] The corresponding rate per carrier is[19,20]

$$\left(\frac{\partial E}{\partial t}\right)_{e-p} = -\frac{(\hbar\omega_p)^2}{a_0 (2m_e^* E)^{1/2}} \ln\left(\frac{E}{\hbar\omega_p}\right)^{1/2}, \tag{3}$$

where a_0 is the Bohr radius. Using the values from above for the parameters, $(\partial E/\partial t)_{e-p}$ is approximately -10^{13} eV/sec. Plasmons can subsequently decay via single-particle excitation (or, if $\hbar\omega_p$ is greater than the energy gap, E_G, by electron-hole pair creation).

For the high carrier densities present ($N > 10^{19}$ cm^{-3}), the plasmon energy $\hbar\omega_p$ is much greater than phonon energies $\hbar\omega$. Consequently, plasmon-phonon coupling is weak and phonon production by the plasmons is negligible. Since no energy is transferred to the lattice by the creation of plasmons and their decay, the energy lost by the hot electron remains in the carrier system. Near the start of the laser pulse, the plasma frequency passes through resonance with the phonons, but the rise in N is so fast that the amount of energy transferred to the phonons in this time is a negligible fraction of the total pulse energy. (With $g \sim 10^{31}$ cm^{-3} sec^{-1}, $\hbar\omega_p$ is resonant with phonon energies for $\sim 10^{-14}$ sec, so that the energy transferred during that time is 5×10^{-3} eV cm^{-3}.)

In addition, the carriers may relax by emitting phonons. Again we emphasize that in this process, as opposed to those relaxation channels discussed above, the energy of the hot carrier is not merely redistributed among the remaining carriers but is transferred from the plasma to the lattice. Under the same conditions for which we have estimated the other rates, we shall find $(\partial E/\partial t)_{e-ph} \sim -10^{11}$ eV/sec. This process will be examined in detail in Sec. V.

Comparing the above results, we find that the rates of energy relaxation by both intercarrier collisions and plasmon production dominate energy relaxation by phonon emission. The reason for this result is that $-(\partial E/\partial t)_{e-e}$ and $-(\partial E/\partial t)_{e-p}$ increase with N_e, whereas, as we shall see in Sec. V, $-(\partial E/\partial t)_{e-ph}$ does not. Clearly, for N_e sufficiently large, the former rates will dominate. This criterion is met for $N_e > 10^{18}$ cm^{-3}, so will be satisfied at densities attained in our system.

Arguments analogous to those relevant to electron energy loss rates lead to similar conclusions about hot holes. Consequently, we expect that collisions between carriers result in their rapid thermalization. In times of order 10^{-14} sec, the electrons and holes attain thermal distributions

characterized by temperatures $T_e = T_h$. Carrier thermalization is achieved without transferring energy to the lattice, so that the carrier temperature is initially much higher than the temperature of the lattice, T_L.

III. CARRIER RECOMBINATION

Because the carriers are thermalized, both electrons and holes can be described by quasi-Fermi levels and a single temperature $T_e = T_h$. In order to determine the relationship between these two levels we must consider recombination. Auger processes, with rates proportional to N^3, will dominate recombination at high carrier densities.[21] Specifically, an electron recombines with a hole; the energy released is taken up by a third carrier. This hot carrier will rapidly thermalize with the rest of the carriers through collisions and plasmon production as discussed in Sec. II. The Auger rate is given by $(\partial N_e/\partial t)_{\text{Auger}} = -C_e N_e^2 N_h - C_h N_h^2 N_e = -(C_e + C_h) N_e^3 = -C N_e^3$. For crystalline Si, $C \approx 4 \times 10^{-31}$ cm^6 sec, where C increases only weakly with temperature.[22]

A sufficiently dense plasma will partially screen the Coulomb interaction between carriers, leading to a modified Auger rate[23]

$$\left(\frac{\partial N_e}{\partial t}\right)_{\text{Auger}} = \frac{-C N_e^3}{[1 + (\lambda/k_G)^2]^2} = -C'(N_e) N_e^3 ,$$

where λ is the screening wave vector, and $k_G = (2m_e^* E_G)^{1/2}/\hbar$, with E_G the energy gap, is the approximate wave vector of the most likely recombination transition. As we shall see in Sec. IV, the high carrier temperatures lead to nondegeneracy even at high densities, so that $\lambda^2 \sim N_e$, and the recombination time $\tau_{\text{Auger}} = N_e/(\partial N_e/\partial t)_{\text{Auger}}$ approaches a constant for very large N_e. This screening becomes important at $N_e \sim 10^{21}$ cm^{-3}, and as N_e further increases, τ_{Auger} decreases asymptotically to approximately 6×10^{-12} sec. Carrier excitation occurs at the same time by the inverse process, impact ionization, with $(\partial N_e/\partial t)_{\text{ioniz}} = -C'(N_e) N_0^2 N_e$, where N_0 is the equilibrium value of N_e at T_e.

During the laser pulse, the rate of change of electron density in the irradiated volume is therefore

$$\frac{dN_e}{dt} = \left[\left(\frac{\partial N_e}{\partial t}\right)_{\text{gen}} + \left(\frac{\partial N_e}{\partial t}\right)_{\text{diff}}\right] + C'(N_e) N_e (N_0^2 - N_e^2) , \tag{4}$$

where $(\partial N_e/\partial t)_{\text{gen}}$ is the electron laser generation rate and $(\partial N_e/\partial t)_{\text{diff}}$ is the rate at which electrons diffuse from this volume. When $N_e = N_h = N_0$, impact ionization balances Auger recombination, so that the electrons and holes are in equilibrium with each other and are described by a common quasi-Fermi level. We want to find the conditions for which departure from this equilibrium is small. N_0 is a function of the carrier energy density, and therefore varies with time as the laser inputs energy, as energy diffuses away, and as phonons are emitted. N_e follows the changes in N_0 provided the net ionization-recombination rate is faster than the remaining terms in Eq. (4). If $N_e \approx N_0$, Eq. (4) becomes

$$\frac{dN_e}{dt} \approx G + 2C'(N_0) N_0^3 - 2C'(N_0) N_0^2 N_e , \tag{5}$$

where $G \equiv (\partial N_e/\partial t)_{\text{gen}} + (\partial N_e/\partial t)_{\text{diff}}$. For N_e sufficiently large, $2C'(N_0) N_0^3 \gg G$. Because $(\partial N_e/\partial t)_{\text{gen}} \leq g$ (due to free carrier absorption) and $(\partial N_e/\partial t)_{\text{diff}} \leq 0$, $G \leq g$. Therefore, when $2C'(N_0) N_0^3 \gg g$, we can certainly ignore the first term in Eq. (5). This condition is satisfied for $N_0 \gtrsim 10^{20}$ cm^{-3}. In this case, $dN_e/dt \approx -2(N_e - N_0)/\tau_{\text{Auger}}$. As a result, if N_0 does not vary appreciably in times of order τ_{Auger}, N_e will be approximately equal to N_0. Since N_0 is determined by the total energy contained in the carriers, $dN_0/dt \leq g$ and variations in N_0 are sufficiently slow to ensure that departures from quasiequilibrium are small.

Because Auger recombination need involve no energy loss from the carrier system, the exact *details* of the absorption of the laser light—i.e., the relative amounts of electron-hole pair excitation and free carrier absorption—are unimportant. The carrier density follows N_0 and is therefore completely determined by the total energy E_{tot} of the carriers.

IV. ENERGY EQUIPARTITION

We can now calculate the dependence of the carrier density and temperature on E_{tot}. N_e equals N_0, so that it obeys the usual expression

$$N_e = \int_0^\infty \frac{g_e(E)\,dE}{1 + e^{(E - E_F)/kT_e}} . \tag{6}$$

At the same time, the single-particle energy of the electrons is

$$E_e = \int_0^\infty \frac{E g_e(E)\,dE}{1 + e^{(E - E_F)/kT_e}} . \tag{7a}$$

In Eqs. (6) and (7a), $g_e(E)$ is the single-particle density of states and E_F is the Fermi energy. Equations analogous to these apply to the holes. In addition, energy is stored in thermally excited collective oscillations of the carriers:

$$E_{\text{coll}} = \left(\frac{k_c}{6\pi^2}\right) \frac{\hbar\omega_p}{e^{\hbar\omega_p/kT_e} - 1} . \tag{7b}$$

$\hbar\omega_p$ is given by Eq. (2). We assume that these

plasmons have negligible k dispersion. $k_c^3/6\pi^2$ is the number of allowed modes at $\hbar\omega_p$. The cutoff wave vector,[13,14] k_c, is taken to be approximately the Debye wave vector,

$$k_c \sim \left(\frac{4\pi N e^2}{\epsilon k T_e}\right)^{1/2}.$$

Assuming ellipsoidal constant energy surfaces characterized by effective masses m_e^* and m_h^*, Eqs. (6) and (7a) become

$$N_e = A_e (m_e^* k T_e)^{3/2} F_{1/2}(\eta), \tag{8}$$

$$E_e = A_e m_e^{*3/2} (k T_e)^{5/2} F_{3/2}(\eta), \tag{9}$$

where

$$A_e \equiv \frac{M_e}{2\pi^2}\left(\frac{2}{\hbar^2}\right)^{3/2},$$

with M_e the number of equivalent conduction-band minima,

$$\eta \equiv \frac{E_F}{k T_e},$$

and

$$F_j(\eta) \equiv \int_0^\infty \frac{x^j\,dx}{e^{x-\eta}+1}$$

is the usual Fermi-Dirac integral. The total energy of the carriers is given by $E_{tot} = N_e E_G + E_e + E_h + E_{e\text{-}h}$. The term $N_e E_G$ represents the energy of excitation across the gap. We have neglected the dependence of E_G on N_e, T_e, and T_L. However, because the carrier temperature T_e is very high, Eqs. (8) and (9) depend only weakly on E_G and the error introduced is not large. In Fig. 1, we have plotted E_{tot} and $N = N_e + N_h$ as functions of carrier temperature kT_e for Si with parameters $M_e = 6$, $M_h = 1$, $m_e^* = 0.33m$, $m_h^* = 0.55m$, and $m_r^* = 0.15m$.[18] As expected, both number and energy are monotonically increasing functions of temperature.

The importance of the plasmons is not readily apparent from Fig. 1. To demonstrate their role more clearly, we look at the dependence of kT_e on $(E_{tot} - N_e E_G)/N$, the excess energy per carrier. Using Eqs. (7)–(9), we find

$$\frac{E_{tot} - N_e E_G}{N} = k T_e \left(\frac{F_{3/2}(\eta)}{F_{1/2}(\eta)}\right) + \frac{e^2}{3\pi\hbar\epsilon}(m_r^* k T_e)^{1/2} \times \frac{z^2}{e^z - 1}, \tag{10}$$

where $z \equiv \hbar\omega_p / kT_e$. kT_e is plotted in Fig. 2 as a function of the per carrier excess energy. For small N, $kT_e \to \frac{2}{3}[(E_{tot} - N_e E_G)/N]$, as expected. The dashed curve was derived by neglecting the plasmon contribution. Note that the temperature falls below this limit when an appreciable amount of energy

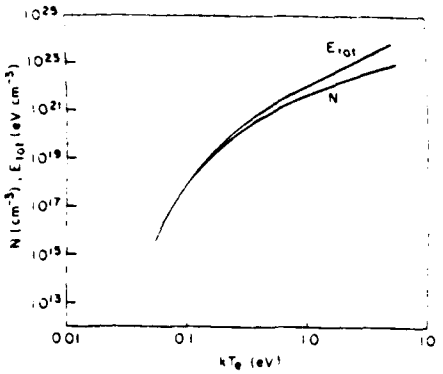


FIG. 1. E_{tot} and N as functions of carrier temperature kT_e calculated for Si with $M_e = 6$, $M_h = 1$, $m_e^* = 0.33m$, $m_h^* = 0.55m$, and $m_r^* = 0.15m$.

is contained in the thermally excited plasma oscillations. This occurs when the carrier temperature and plasmon energies are comparable. In Fig. 3, we plot the fraction of the excess energy contained in the plasmons as a function of carrier temperature. At the peak of the curve, roughly ten per cent of the energy is partitioned into the plasmons. This particular value should not be taken too seriously, however, because it depends critically on our estimate of k_c.

For typical doped semiconductors $N_e \sim 10^{17}$–10^{19}

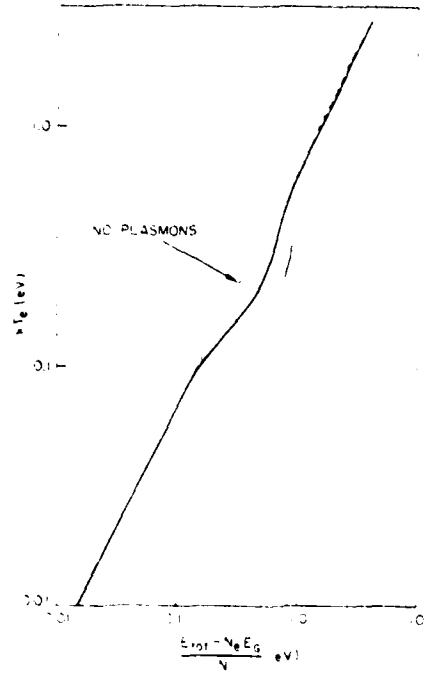


FIG. 2. kT_e as a function of $(E_{tot} - N_e E_G)/N$. The dashed curve was derived by neglecting the plasmon contribution.

cm^{-3}) at room temperature, $\hbar\omega_p \sim 10^{-4}$–$10^{-2}$ eV $\ll kT$. As a result, many collective modes may be occupied, but their energies are so small they contain a very small fraction of the total energy. For metals, on the other hand, $\hbar\omega_p \sim 10$ eV $\gg kT$. The plasmon energies are very large, but so large that a negligible number of them are thermally excited. It is at just those temperatures ($kT_e \sim 0.1$–1 eV) and densities ($N \sim 10^{19}$–10^{21} cm^{-3}) relevant to our discussion that a significant fraction of the total energy is contained in thermally excited plasma oscillations. The heat capacity of the carrier system is increased as a result of the additional degrees of freedom, into which a significant fraction of the energy is partitioned. In other words, if we try to heat up the carriers by pumping energy into them, their temperature will be lower than in the absence of plasmons, because these modes are absorbing some of the energy.

V. PHONON EMISSION

In this section, we shall first calculate the excitation spectrum for the carrier system and then relate the result to the phonon emission rate. We shall examine this rate in detail, focusing on emission at those $\vec{q}$'s most relevant to silicon, and paying particular attention to the effects of the hot, dense carriers.

We begin by calculating the contribution to the imaginary part of the dielectric function, $\epsilon_2(\vec{q},\omega)$, from electronic transitions within and between conduction-band minima. We do not sum over valence to conduction-band transitions. The ω's at which we ultimately evaluate the dielectric function are far from resonance with interband excitation energies so that these transitions do not contribute significantly to ϵ_2. In the random-phase approximation (RPA),[25]

$$\epsilon_2(\vec{q},\omega) = \frac{4\pi^2 e^2}{\Omega q^2} \sum_{ij} \sum_{\vec{k}\sigma} f^i_{\vec{k}\sigma}(1 - f^j_{\vec{k}-\vec{q},\sigma}) \times \delta(\hbar\omega - [\Delta E(\vec{k},\vec{q})]_{ij}), \tag{11}$$

where Ω is the volume. The sum $\sum_{ij}$ ranges over all pairs of valleys i and j (including $i=j$) where the transitions are from valley i to valley j. The occupation probability of a state $\vec{k}\sigma$ in the ith valley is $f^i_{\vec{k}\sigma}$ and $[\Delta E(\vec{k},\vec{q})]_{ij}$ is the difference in energy between electronic states $(\vec{k}+\vec{q},\sigma)_j$ and $(\vec{k}\sigma)_i$. In RPA, the electrons have free-electron-like wave functions and polarizabilities, but respond to an effective Coulomb potential which includes screening self-consistently, in contrast to the Hartree-Fock approximation, in which the electrons respond to just the external field.[25]

We approximate each conduction-band minimum by a spherical constant-energy surface having effective mass m_e^*. The wave vector from a particular valley to its jth neighbor is denoted by $\vec{Q}_{oj}$. In this case, we find the following equivalence (see Fig. 4): For any *intervalley* transition $\vec{k} \to \vec{k}+\vec{q}$, there is a corresponding *intravalley* transition $\vec{k} \to \vec{k}+\vec{Q}_j$, where $\vec{Q}_j \equiv \vec{q} - \vec{Q}_{oj}$, such that (a) $f^i_{\vec{k}\cdot\vec{Q}_j} = f^i_{\vec{k}+\vec{q}}$, and (b) $E^i_{\vec{k}\cdot\vec{Q}_j} = E^i_{\vec{k}+\vec{q}}$. Within our approximation for the energy surfaces, the results (a) and (b) follow directly from the definition of $\vec{Q}_j$. [Actually, this equivalence holds exactly in Si for those pairs of *ellipsoidal* valleys aligned on a common ⟨1,0,0⟩ axis.] Equation (11) therefore becomes

$$\epsilon_2(\vec{q},\omega) = \frac{8\pi^2 e^2}{\Omega q^2} \sum_{ij} \sum_{\vec{k}} f^i_{\vec{k}}(1 - f^j_{\vec{k}\cdot\vec{Q}_j}) \delta[\hbar\omega - (E_{\vec{k}\cdot\vec{Q}_j} - E^i_{\vec{k}})], \tag{12}$$

where we have included the intravalley transition by defining $\vec{Q}_{jj,\,\mathrm{intra}} \equiv 0$ so that $\vec{Q}_{j,\,\mathrm{intra}} = \vec{q}$. In Eq. (12), $E_{\vec{k}\cdot\vec{Q}_j} - E^i_{\vec{k}} = \hbar^2(Q_j^2 + 2kQ_j\cos\Theta)/2m_e^*$, where Θ is the angle between $\vec{k}$ and $\vec{Q}_j$. The important simplification we have made is that we can now include *intervalley* transitions by evaluating the sums within any *one* valley.

The high temperatures ensure that the occupation probabilities are essentially Boltzmann, so that $f^i_{\vec{k}} \sim e^{-(E^i_{\vec{k}} - E_F)/kT_e}$ and $f(1-f) \sim f$. Making these substitutions, we obtain

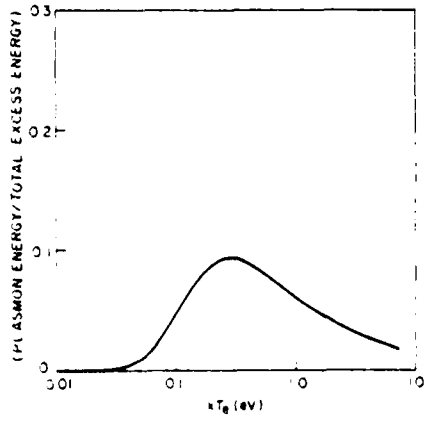


FIG. 3. Fraction of excess carrier energy contained in the plasmons as a function of carrier-temperature kT_e.

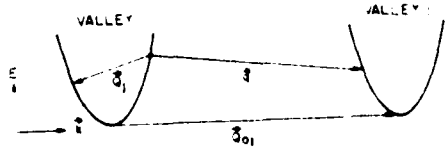


FIG. 4. Schematic of intervalley-intravalley equivalence.

$$\epsilon_2(\vec{q},\omega)=\frac{8\pi^2e^2}{\Omega q^2}e^{-(E_c-E_F)/kT_e}\sum_i\sum_{\vec{k}}e^{-\hbar^2k^2/2m_e^*kT_e}\,\delta\left(\hbar\omega-\frac{\hbar^2}{2m_e^*}(2kQ_j\cos\Theta+Q_j^2)\right), \tag{13}$$

where E_c is the energy of the conduction-band minimum. When we let $\sum_{\vec{k}}=(\Omega/8\pi^3)\int d^3k$, sum over i, and integrate over the δ function, Eq. (13) becomes

$$\epsilon_2(\vec{q},\omega)=\frac{M_e e^2}{2q^2}e^{-(E_c-E_F)/kT_e}\sum_j\int_{k_{min}}^{\infty}\frac{k^2\exp\left(-\frac{\hbar^2k^2}{2m_e^*kT_e}\right)dk}{\left|\frac{\hbar^2kQ_j}{2m_e^*}\right|}, \tag{14}$$

where

$$k_{min}=\left|\frac{m_e^*}{\hbar^2Q_j}\left(\hbar\omega-\frac{\hbar^2q^2}{2m_e^*}\right)\right|.$$

Integrating Eq. (14), we find

$$\epsilon_2(\vec{q},\omega)=\frac{(8\pi)^{1/2}(m_e^*kT_e)^{3/2}M_e}{\hbar^2q^2}e^{-(E_c-E_F)/kT_e}\sum_j\left(\frac{Z_je^{-(\hbar\omega-E_{Q_j})^2/4kT_eE_{Q_j}}}{(4\pi kT_eE_{Q_j})^{1/2}}\right), \tag{15}$$

where $E_{Q_j}\equiv\hbar^2Q_j^2/2m_e^*$, Z_j is the number of valleys separated from valley i by Q_{0j}, and the sum is over normalized Gaussians weighted by Z_j. Finally, using the relation

$$N_e=2M_e\left(\frac{2\pi m_e^*kT_e}{h^2}\right)^{3/2}e^{-(E_c-E_F)/kT_e},$$

the imaginary part of the dielectric function becomes

$$\epsilon_2(\vec{q},\omega)=\frac{4\pi^2N_ee^2}{q^2}\sum_j\frac{Z_je^{-(\hbar\omega-E_{Q_j})^2/4kT_eE_{Q_j}}}{(4\pi kT_eE_{Q_j})^{1/2}}. \tag{16}$$

In Fig. 5, we sketch $\epsilon_2(\vec{q},\omega)$ as a function of $\hbar\omega$ for an arbitrary fixed $\vec{q}$. Note that there are Gaussian peaks centered at $\hbar\omega_j=E_{Q_j}$. These peaks correspond to transitions originating from the densely populated states near the band minima. The widths of the peaks are $(2kT_eE_{Q_j})^{1/2}$. At higher temperatures the peaks are broader because electrons occupy a larger volume in k space, thereby increasing the likelihood of transitions originating from higher in the bands. Of course, the peak heights are complementary to their widths—those resonances that are more spread out are less strong at any given ω.

To put $\epsilon_2(\vec{q},\omega)$ in a more workable form, we approximate the normalized Gaussians by normalized Lorentzians having the same peak locations and widths. Then, using the Kramers-Kronig relations we can obtain $\epsilon_1(\vec{q},\omega)$, the real part of the dielectric function, and, consequently, $\epsilon(\vec{q},\omega)=\epsilon_1(\vec{q},\omega)+i\epsilon_2(\vec{q},\omega)$. This function is

$$\epsilon(\vec{q},\omega)=\epsilon_1(\vec{q},\omega)+\frac{4\pi N_ee^2}{m_e^*}\sum_jZ_j\left(\frac{Q_x}{q}\right)^2\left(\frac{\omega_j^2-\omega^2}{(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2}+\frac{i\Gamma_j\omega}{(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2}\right), \tag{17}$$

where $\hbar\Gamma_j=(8kT_eE_{Q_j})^{1/2}$. $\epsilon_1(\vec{q},\omega)$ represents the contribution from interband transitions to $\epsilon(\vec{q},\omega)$. As discussed earlier, $\mathrm{Im}\,\epsilon_1(\vec{q},\omega)$ is negligible at those frequencies with which we shall be concerned. Because electron-electron collisions are dominant, collisional broadening widths $\hbar\Gamma_{coll}$ can be evaluated from Sec. II to be ≈0.05 eV. On the other hand, $\hbar\Gamma_j$ ranges from ≈0.2 eV for intravalley transitions to ≈3 eV for transitions between valleys, so that the collisional widths are very small in comparison and we can ignore them.

Next, we calculate the excitation spectrum

$$\mathrm{Im}\,\frac{1}{\epsilon(\vec{q},\omega)}=\frac{\epsilon_2}{\epsilon_1^2+\epsilon_2^2}.$$

When we square ϵ_1 and ϵ_2, we find that we can ignore cross terms for differing Q_j, which are negligible (peak times tail) compared to the terms corresponding to the square of one peak. The excitation spectrum is given by

$$\operatorname{Im}\frac{1}{\epsilon(\vec{q},\omega)}=\frac{1}{\epsilon_0}\left[(\hbar\omega_{pe})^2\sum_j Z_j\left(\frac{Q_j}{q}\right)^2\frac{(\hbar\Gamma_j)(\hbar\omega)}{[E_{Q_j}^2-(\hbar\omega)^2]^2+(\hbar\omega)^2(\hbar\Gamma_j)^2}\right]$$
$$\times\left[1+2(\hbar\omega_{pe})^2\sum_j Z_j\left(\frac{Q_j}{q}\right)^2\frac{E_{Q_j}^2-(\hbar\omega)^2}{[E_{Q_j}^2-(\hbar\omega)^2]^2+(\hbar\omega)^2(\hbar\Gamma_j)^2}\right.$$
$$\left.+(\hbar\omega_{pe})^4\sum_j Z_j^2\left(\frac{Q_j}{q}\right)^4\frac{1}{[E_{Q_j}^2-(\hbar\omega)^2]^2+(\hbar\omega)^2(\hbar\Gamma_j)^2}\right]^{-1}, \tag{18}$$

where $\epsilon_0\equiv\epsilon_0(\vec{q},\omega\approx 0)$ and

$$\omega_{pe}\equiv\left(\frac{4\pi N_e e^2}{\epsilon_0 m_e^*}\right)^{1/2}.$$

[ω_{pe} is used for notational convenience and does not correspond to the plasma frequency which is given by Eq. (2).] Earlier, we saw that $\epsilon_2(\vec{q},\omega)$, for fixed $\vec{q}$, was peaked at several values of ω. We now look at Eq. (18) for fixed ω. We find that it is peaked at several q—those q such that $E_{Q_j}\approx\hbar\omega$. (If we had chosen to include hole transitions, the result would simply be the addition here of another peak at $\hbar^2Q_h^2/2m_h^*=\hbar\omega$. However, because all the hole transitions are intravalley, $\vec{Q}_h=\vec{q}$. Since we shall be looking at small $\hbar\omega$, phonon emission by holes therefore occurs primarily at small q. As we shall see, these transitions are well screened even at moderate carrier densities. This will also be the case for intravalley electron transitions, which we can take to be representative of both types of small-q phonon emission.)

In general, each intervalley separation $\vec{Q}_{0_j}$ will have a corresponding $\vec{q}$ such that $E_{Q_j}\approx\hbar\omega$. $\operatorname{Im}(1/\epsilon)$ is peaked at these values of $\vec{q}$, which we denote by $\vec{q}_j$. Near these peaks,

$$\operatorname{Im}\frac{1}{\epsilon(\vec{q}_j,\omega)}=\frac{1}{\epsilon_0}\frac{(\hbar\omega_{pe})^2Z_j(Q_j/q_j)^2\hbar\omega(8kT_e\hbar\omega)^{1/2}}{8kT_e(\hbar\omega)^3+(\hbar\omega_{pe})^4Z_j^2(Q_j/q_j)^4}. \tag{19}$$

The value of $\vec{q}_j$ at each of the peaks corresponds to a particular intervalley transition and, for small ω, $\vec{q}_j\approx\vec{Q}_{0_j}$. The widths of these peaks do not vary with q_j, since $\hbar\Gamma_j=(8kT_eE_{Q_j})^{1/2}\approx(8kT_e\hbar\omega)^{1/2}\neq f(q_j)$ for fixed ω.

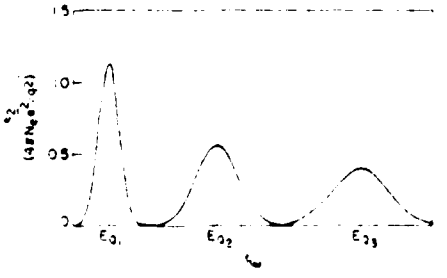


FIG. 5. $\epsilon_2/(4\pi^2N_e e^2/q^2)$ as a function of $\hbar\omega$, for arbitrary fixed $\vec{q}$.

We now relate the excitation spectrum to the phonon emission rate[25]

$$\frac{dN_{\vec{q},\omega}}{dt}=\frac{2\pi}{\hbar}\sum_{ij}\sum_{\vec{k},\sigma}\left|\frac{V_{\vec{k}+\vec{q},\vec{k}}}{\epsilon}\right|^2 f_{\vec{k}+\vec{q},\sigma}^{i}(1-f_{\vec{k}\sigma}^{j})$$
$$\times\delta(\hbar\omega-(E_{\vec{k}+\vec{q}}^{i}-E_{\vec{k}}^{j})), \tag{20}$$

where $V_{\vec{k}+\vec{q},\vec{k}}$ is the unscreened matrix element for the transition of an electron from state $\vec{k}+\vec{q}$ to state $\vec{k}$, with the emission of a phonon of wave vector $\vec{q}$. As is usual, we assume V is not a function of the electronic state $\vec{k}$.[26] Evaluating Eq. (20) in the same manner as we did Eq. (12), we find

$$\frac{dN_{\vec{q},\omega}}{dt}=\frac{2\pi}{\hbar}N_e\left|\frac{V}{\epsilon}\right|^2 e^{-\hbar\omega/kT_e}\sum_j\frac{Z_j e^{-(\hbar\omega-E_{Q_j})^2/4kT_eE_{Q_j}}}{(4\pi kT_eE_{Q_j})^{1/2}}. \tag{21}$$

Comparing this result with Eq. (16) and using Eq. (17), we see that

$$\frac{dN_{\vec{q},\omega}}{dt}=\frac{m_e^*V^2E_{\vec{q}}}{\pi e^2\hbar^2}\left(\operatorname{Im}\frac{1}{\epsilon(\vec{q},\omega)}\right)e^{-\hbar\omega/kT_e}. \tag{22}$$

The factor $e^{-\hbar\omega/kT_e}$ appears for phonon emission because carriers must be an energy $\hbar\omega$ higher in the band than required for the absorption transition that contributes to ϵ_2. With the use of Eq. (19), the emission rate can be written as

$$\frac{dN_{\vec{q}_j,\omega}}{dt}=\frac{m_e^*V^2}{\epsilon_0\pi e^2\hbar^2}$$
$$\times e^{-\hbar\omega/kT_e}\frac{(\hbar\omega_{pe})^2Z_j(\hbar\omega)^2(8kT_e\hbar\omega)^{1/2}}{(\hbar\omega)^3(8kT_e)+(\hbar\omega_{pe})^4Z_j^2(\hbar\omega-E_{i,j})^4}. \tag{23}$$

Because $\hbar\omega\ll E_{Q_{0_j}}$ for intervalley transitions, $\vec{q}_j\approx\vec{Q}_{0_j}$. (For transitions within a valley, $\vec{Q}_{i,j}=0$, so we treat $\vec{q}_j$ explicitly, with $E_{i,j}\approx\hbar\omega$.) As a result, phonon emission is peaked at those phonons with wave vectors corresponding to the intervalley separations.

$$\frac{dN_{\vec{q}_j,\omega}}{dt}\approx\frac{dN_{\vec{Q}_{0_j},\omega}}{dt}$$
$$=\frac{(\hbar\omega_{pe})^2Z_j(m_e^*V^2/\epsilon_0\pi e^2\hbar^2)e^{-\hbar\omega/kT_e}}{(8kT_e\hbar\omega)^{1/2}[1+(\hbar\omega_{pe})^4Z_j^2/8kT_e\hbar\omega E_{i,j}^4]}. \tag{24}$$

In Fig. 6, we sketch $(1/N_e)(dN_{\vec{q}_j,\omega}/dt)$, the rate of phonon emission per electron, as a function of

$(\hbar\omega_{pe})^2$, where we focus on phonons having wave vector $\vec{Q}_{0_j}$ and frequency ω. Notice the abrupt change in behavior at

$$(\hbar\omega_{pe})^2 = (\hbar\omega_{pe})^2_{crit} \approx \left(\frac{(\hbar\omega)(EQ_{0_j})^2 8kT_e}{Z_j^2}\right)^{1/2}. \quad (25)$$

Since $(\hbar\omega_{pe})^2$ is proportional to N_e, this means that below a critical density of electrons, the per electron emission rate is independent of N_e. In this regime

$$\frac{1}{N_e}\frac{dN_{\vec{q}_j,\omega}}{dt} = \frac{2Z_j V^2 e^{-\hbar\omega/kT_e}}{\hbar\epsilon_1 [2kT_e(\hbar\omega)]^{1/2}}. \quad (26)$$

However, for sufficiently large N_e, screening by these carriers becomes important and the rate falls rapidly with N_e:

$$\frac{1}{N_e}\frac{dN_{\vec{q}_j,\omega}}{dt} = \frac{m_e^* V^2 E_{Q_{0_j}}^2 [8kT_e(\hbar\omega)]^{1/2} e^{-\hbar\omega/kT_e}}{\epsilon_1 Z_j \pi e^2 \hbar^3}\left(\frac{1}{(\hbar\omega_{pe})^2}\right). \quad (27)$$

If $N_e(T_{e1})$ is much less than the critical value given by Eq. (25), then

$$\frac{\left(\frac{1}{N_e}\frac{dN_{\vec{q}_j,\omega}}{dt}\right)_{T_{e2}}}{\left(\frac{1}{N_e}\frac{dN_{\vec{q}_j,\omega}}{dt}\right)_{T_{e1}}} = \frac{\left(\frac{T_{e1}}{T_{e2}}\right)^{1/2} e^{\hbar\omega(1/kT_{e1}-1/kT_{e2})}}{1+\frac{[\hbar\omega_{pe}(T_{e2})]^4 Z_j^2}{8kT_{e2}(\hbar\omega)(E_{Q_{0_j}})^2}}. \quad (28)$$

We see that for $T_{e2} > T_{e1}$, screening by the large number of excited carriers increases the phonon emission time considerably from its unscreened value.

Although the dominant energy loss mechanism is phonon emission by electrons (hole transitions are effectively screened), rapid electron-hole equilibration ensures that both species lose energy at the same per carrier rate,

$$\left(\frac{\partial E}{\partial t}\right)_{e\text{-ph}} = -\frac{\hbar\omega}{N}\sum_j \frac{dN_{\vec{q}_j,\omega}}{dt} \sim -\frac{\hbar\omega}{\tau}. \quad (29)$$

In general, phonon absorption occurs in addition to phonon emission. The net rate of energy relaxation therefore depends on both carrier and lattice temperatures, through terms such as[32] $e^{-\hbar\omega/kT_e} - e^{-\hbar\omega/kT_L}$. In our case, $T_e \gg T_L$, so that there is no explicit dependence on lattice temperature and only phonon emission is important.

We can now evaluate Eq. (25) to find the density at which screening effects become important. Assuming intervalley separations appropriate to crystalline Si,[27-29] there are three classes of transitions for the electrons: (a) Intravalley transitions are represented by $\vec{Q}_0 = 0$ and $Z = 1$. We shall consider phonon emission at $\hbar\omega \approx 0.05$ eV, leading to a corresponding $q = 6.6 \times 10^7$ cm^{-1}. (b) All transitions from one valley to another on the same axis can be reduced to one $\vec{Q}_0$, e.g., $\vec{Q}_0 = (2\pi/a)(0,4,0,0)$, where a is the lattice parameter, with $Z = 1$, and $q \approx 4.6 \times 10^7$ cm^{-1}. (c) Finally, there are four inequivalent transitions ($Z = 4$) within the first zone from one valley to another on a different axis, e.g., $\vec{Q}_0 = (2\pi/a)(-,2,1,\pm,2)$ and $(2\pi/a)(-,2,\pm,2,1)$, so that $q \approx 1.2 \times 10^8$ cm^{-1}. For $\epsilon_1(\vec{q},\omega)$ (implicit in ω_{pe}) we use the values calculated by Walter and Cohen.[30] [Although their calculations were performed for $\vec{q}$ along the (1,0,0) direction, and therefore apply strictly only to scattering between valleys on the same axis, we use their results as an estimate of ϵ_1 for off-axis scattering also.]

As a result, Eq. (25) becomes

(a) $(\hbar\omega_{pe})^2_{crit} \approx 0.03(kT_e)^{1/2}$,

(b) $(\hbar\omega_{pe})^2_{crit} \approx 0.85(kT_e)^{1/2}$,

(c) $(\hbar\omega_{pe})^2_{crit} \approx 0.53(kT_e)^{1/2}$,

where the energies are in eV. Using the relationship between N_e and T_e found in Sec. IV, we arrive at the critical densities at which screening begins to become important for (a) intravalley, (b) intervalley, on-axis, (c) intervalley, off-axis scattering, respectively,

(a) $N_{e,crit} \approx 2.5 \times 10^{19}$ cm^{-3},

(b) $N_{e,crit} \approx 1.9 \times 10^{21}$ cm^{-3},

(c) $N_{e,crit} \approx 1.0 \times 10^{21}$ cm^{-3}.

The rates in the strong-screening regime are proportional to $1/N_e^2$, so that, e.g., a factor of 5 increase in N_e reduces the rates by 25. Unscreened energy loss rates are reported to be[24] $\sim 10^{11}$ eV/sec; from Eq. (24) we see that for $N_e \approx 5 \times 10^{21}$ cm^{-3}, the rate may be reduced to $\sim 10^{10}$ eV/sec. We shall demonstrate in the following section, however, that carrier densities attained

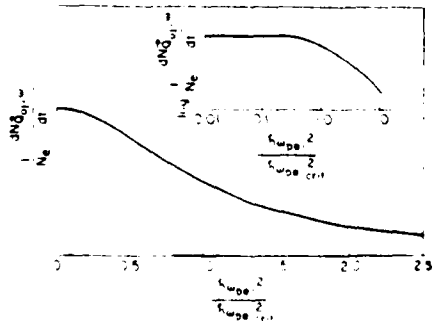


FIG. 6. $(1/N_e)dN_{\vec{Q}_0,j,\omega}/dt$ as a function of $(\hbar\omega_{pe})^2/(\hbar\omega_{pe})^2_{crit}$. The inset shows $\log[1/N_e(dN_e/dt)]$ vs $\log[(\hbar\omega_{pe})^2/(\hbar\omega_{pe})^2_{crit}]$.

are unlikely to exceed the critical densities for screening of intervalley transitions (although screening of intravalley transitions will occur). A reliable estimate of the phonon emission rate depends critically on the electron band structure, which is difficult to determine under extreme conditions of high implantation dosages, amorphization, and rapid atomic rearrangement. Of course, in all of the above calculations we have assumed that the concept of k space is valid and we have evaluated the resulting expressions at those particular values of $\vec{q}$ appropriate to crystalline Si. The strict applicability of this approach depends on the degree of amorphization which itself varies rapidly as a result of the annealing process. As pointed out by Dumke,[31] it may be that transition rates in the noncrystalline state are at least as fast as in the crystal, owing to partial relaxation of selection rules.

VI. CARRIER DIFFUSION

Most of the laser energy is absorbed by the carriers within an absorption depth δ. Eventually, these carriers lose their energy to the lattice: the rise in lattice temperature depends on the distance they have diffused before substantial phonon emission occurs. Now that we know how N_e varies with E_{tot} and the dependence of the phonon emission rate on N_e and T_e, we can examine the manner in which carrier diffusion redistributes the energy. In particular, we shall calculate the effect of diffusion on the rate of phonon emission in a volume near the surface. It is important to emphasize that this is not a simple diffusion process. Although energy is a conserved quantity, the number of carriers is not. When a particular carrier diffuses, it takes its energy with it, thereby reducing the total energy left behind and consequently lowering the carrier temperature. As we have shown in Sec. III, the rapid rate of recombination ensures that N_e is completely determined as a function of time by this temperature. Therefore, N_e decreases not only as a direct result of carrier diffusion but also as an indirect result of the accompanying energy diffusion.

Because of the strong coupling between the electrons and the holes, they must diffuse at the same rate, which is given by

$$\left(\frac{\partial N}{\partial t}\right)_{diff} = D_a \frac{\partial^2 N}{\partial x^2}, \tag{30}$$

where $D_a = 2D_e D_h / (D_e + D_h)$ is the ambipolar diffusion coefficient. We can estimate D_a using the Einstein relation

$$D_a = \left(\frac{kT_e}{e}\right)\left[\frac{2\mu_e\mu_h}{(\mu_e + \mu_h)}\right],$$

where μ_e (μ_h) is the electron (hole) mobility determined by the carrier-phonon scattering time.[32] This approximation should yield a reasonable estimate even when carrier-carrier collisions are the dominant interaction, as these collisions do not affect the net plasma momentum. $\mu = e\tau/m^*$, so that $D_a = 2kT_e\tau_e\tau_h/(m_e^*\tau_h + m_h^*\tau_e)$. D_a is therefore $\sim 10^2$ cm^2 sec. Because of the exponential dependence of N on T_e, diffusion terms involving $T_e \partial N/\partial x$ are larger than those involving $N \partial T_e/\partial x$ by factors $\sim E_G/2kT_e$, so that we are justified in approximating the diffusion by the simple expression Eq. (30).

As a result of carrier diffusion, the carrier energy changes at a rate given by

$$\frac{E_{tot}}{N}\left(\frac{\partial N}{\partial t}\right)_{diff} = \frac{E_{tot}D_a}{N}\frac{\partial^2 N}{\partial x^2}. \tag{31}$$

The equation governing the total carrier energy is then given by

$$\frac{\partial E_{tot}}{\partial t} = g\hbar\omega_L e^{-x/\delta} + \frac{D_a E_{tot}}{N}\frac{\partial^2 N_e}{\partial x^2} - \frac{N_e\hbar\omega}{\tau}. \tag{32}$$

The last term represents the rate at which phonons with energy $\hbar\omega$ are emitted. Because $E_{tot}/N \equiv \langle E_{tot}\rangle$ and τ are both only weak functions of N_e, the steady-state value of N_e, $N_{e,ss}$ is

$$N_{e,ss}(x) = \frac{g\hbar\omega_L\tau}{\hbar\omega}\frac{\delta\alpha}{1-\delta^2\alpha^2}\left(e^{-x/\alpha} - \frac{\delta}{\alpha}e^{-x/\delta}\right), \tag{33}$$

where $\alpha \equiv (D_a\tau\langle E_{tot}/\hbar\omega\rangle)^{1/2}$. To obtain this solution, we have used energy conservation via the relation

$$\int_0^\infty g\hbar\omega_L e^{-x/\delta}dx = \int_0^\infty \frac{N_{e,ss}(x)\hbar\omega}{\tau}dx. \tag{34}$$

At the surface,

$$N_{e,ss}(x=0) \equiv N_{e,ss}^0 = \frac{g\hbar\omega_L\tau}{\hbar\omega(\alpha/\delta+1)},$$

which we can rewrite in the form

$$\frac{N_{e,ss}^0\hbar\omega}{\tau} = \frac{g\hbar\omega_L}{\alpha/\delta+1}. \tag{35}$$

Equation (35) directly relates the rate of energy loss by phonon emission to the rate of energy input by the laser. In the absence of diffusion ($\alpha = 0$), the two rates are equal. However, we shall find for our example that α and δ are comparable, so that diffusion reduces the rate of energy transfer to the lattice in the region near the surface.[34]

VII. DISCUSSION

We have evaluated $(\partial E/\partial t)_{e\text{-}ph}$ numerically by using Eq. (28) along with a moderate-density

phonon scattering time $\tau \sim 10^{-13}$ sec which is consistent with luminescence data for Si analyzed by Folland[28] and with those values inferred from transport measurements.[35] kT_e and E_{tot} were calculated self-consistently at each N_e as discussed in Sec. IV. Screening does not affect the rate of intervalley phonon emission until $N_e \sim 10^{21}$ cm^{-3}, at which point this rate begins to deviate from its linear N_e dependence. Phonon emission for intravalley transitions falls off at smaller values of N_e. Because screening increases the electron-phonon scattering time, it not only decreases the rate of phonon emission but in addition enhances diffusion. Although for the specific example we are discussing $N^0_{e,ss} < 10^{21}$ cm^{-3} and therefore screening effects do not play a major role in the rate of energy transfer from the carriers to the lattice, at higher excitation rates these effects can be very important. In Fig. 7 we plot $N^0_{e,ss}$ as a function of $g\hbar\omega_L$ (cf. Eq. (36)) for the example described in Sec. I. In this case, $g\hbar\omega_L = P(1-R)\ 5 \approx 3 \times 10^{21}$ eV cm^{-2} sec, so that $N^0_{ss} = 2N^0_{e,ss} \sim 10^{20}$ cm^{-3}, which is the value we have used for numerical estimates throughout this paper. The fact that $N^0_{e,ss}$ has turned out to be larger than 10^{19} cm^{-3} justifies our original assumption that the carrier densities are so high that collisions between carriers are the dominant interactions, that the plasma frequency is much larger than phonon frequencies, and that Auger recombination is most important. The dashed line in the figure indicates values of $N^0_{e,ss}$ obtained by neglecting carrier diffusion ($\alpha = 0$). By comparing the two curves, we see that at a given laser power, diffusion decreases the carrier density near the surface. The rate of energy transfer to the lattice in that region is consequently reduced. Excited carriers may therefore diffuse an appreciable distance from the surface before they give their energy to the lattice, thereby increasing the volume of the region in which this energy transfer occurs.

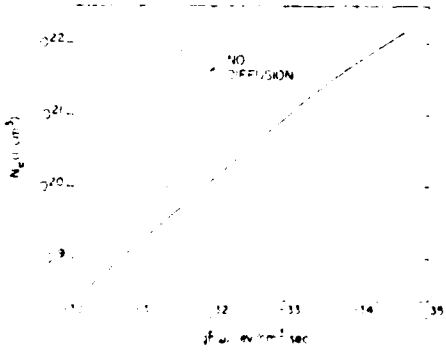


FIG. 7. $N^0_{e,ss}$ as a function of $g\hbar\omega_L$. The dashed curve was derived by neglecting carrier diffusion.

At high excitation rates, the laser energy is given to the lattice within a characteristic depth determined primarily by carrier diffusion. Owing to the extreme nonlinearity of the hot-carrier effects, the uncertainty in the parameters chosen for our example along with the simplifying assumptions made in our calculations prevents us from making an accurate estimate of the precise temperature to which the lattice is heated or from determining the laser power threshold above which melting will occur. However, because the ultimate temperature rise depends so strongly on the extent of the region in which the energy transfer takes place, carrier diffusion plays a significant role in determining this final temperature.

In summary, we have found that for a photon absorption rate $g \sim 10^{21}$ cm^{-3} sec^{-1} at laser wavelength $\lambda_L = 0.53$ μm and pulse length $\tau_L \sim 10$ nsec, plasmon emission by very hot carriers dominates phonon emission, and collisions between these carriers are so rapid that electrons and holes thermalize in times $\sim 10^{-14}$ sec. Auger recombination, in which the energy remains in the carrier system, is the dominant recombination mechanism at these densities. In times $\sim 10^{-13}$ sec, the electrons and holes recombine, reaching a concentration which then follows changes in the plasma energy. A fraction of the carrier energy is partitioned into thermally excited plasmons, thereby raising the heat capacity of the plasma and decreasing its temperature. By equating the rate at which the energy is given to the plasma to the rate at which energy leaves the carriers via diffusion and phonon emission, we can find the steady-state electron density. For our example, $N^0_{e,ss} \sim 10^{20}$ cm^{-3}. At this density, only intravalley transitions are effectively screened. Carrier diffusion plays an important role in determining the ultimate extent to which the lattice is heated.

ACKNOWLEDGMENTS

This work was supported in part by the Air Force Office of Scientific Research under Contract No. F49620-77-C-0005. The author is especially indebted to J. A. Van Vechten for suggesting this problem as well as for many stimulating discussions and continual encouragement. She wishes to thank R. C. Frye, W. P. Dumke, P. J. Price, and F. Stern for important suggestions at numerous stages of this work. In addition, the author is grateful to R. Tsu, R. T. Hodgson, R. A. Ghez, A. A. Lucas, and M. H. Brodsky for providing helpful information in their respective areas of expertise.

[1]I. B. Khaibullin, E. I. Shtyrkov, M. M. Zaripov, R. M. Bayazitov, and M. F. Galjautdinov, Radiat. Eff. 36, 225 (1978).
[2]S. I. Romanov, G. A. Kachurin, L. S. Smirov, I. B. Khaibullin, E. I. Shtyrkov, and R. M. Bayazitov, International Conference on Ion Beam Modification of Materials, Paper C12, Budapest, 1978 (unpublished).
[3]J. A. Van Vechten, R. Tsu, F. W. Saris, and D. Hoonhout, Phys. Lett. 74A, 417 (1979); J. A. Van Vechten, R. Tsu, and F. W. Saris, Phys. Lett. 74A, 422 (1979).
[4]R. A. Ghez and R. A. Laff, J. Appl. Phys. 46, 2103 (1975).
[5]J. C. Wang, R. F. Wood, and P. P. Pronko, Appl. Phys. Lett. 33, 455 (1978).
[6]M. Lax, Appl. Phys. Lett. 33, 786 (1978).
[7]J. C. Schultz and R. J. Collins, Appl. Phys. Lett. 34, 34 (1979).
[8]A. Lietoila and J. F. Gibbons, Appl. Phys. Lett. 34, 335 (1979).
[9]A. Elci, M. O. Scully, A. L. Smirl, and J. C. Matter, Phys. Rev. B 16, 191 (1977).
[10]A. Elci, A. L. Smirl, C. Y. Leung, and M. O. Scully, Solid-State Electron. 21, 151 (1978).
[11]J. Shah and R. C. C. Leite, Phys. Rev. Lett. 22, 1304 (1969); J. Shah, Phys. Rev. B 10, 3697 (1974); J. Shah, Solid-State Electron. 21, 43 (1978).
[12]R. Ulbrich, Phys. Rev. B 8, 5719 (1973); R. G. Ulbrich, Solid-State Electron. 21, 51 (1978).
[13]D. H. Auston, S. McAfee, C. V. Shank, E. P. Ippen, and O. Teschke, Solid-State Electron. 21, 147 (1978).
[14]C. V. Shank, R. L. Fork, R. F. Leheny, and J. Shah, Phys. Rev. Lett. 42, 112 (1979).
[15]D. von der Linde and R. Lambrich, Phys. Rev. Lett. 42, 1090 (1979).
[16]R. Tsu, J. E. Baglin, T. Y. Tan, M. Y. Tsai, K. C Park, and R. Hodgson, *Laser-Solid Interactions and Laser Processing—1978*, Materials Research Society, Boston, AIP Conference Proceedings No. 50, edited by S. D. Ferris, H. J. Leamy, and J. M. Poate (AIP, New York, 1979), p. 344.
[17]C. J. Hearn, Proc. Phys. Soc. 86, 881 (1965).
[18]J. E. Kardontchik and E. Cohen, Phys. Rev. Lett. 42, 669 (1979).
[19]D. Pines and D. Bohm, Phys. Rev. 85, 338 (1952).
[20]J. J. Quinn, Phys. Rev. 126, 1453 (1962).
[21]A. R. Beattie and P. T. Landsberg, Proc. R. Soc. London, Ser. A 249, 16 (1959).
[22]J. Dziewior and W. Schmid, Appl. Phys. Lett. 31, 346 (1977).
[23]A. Haug, Solid-State Electron. 21, 1281 (1978).
[24]D. Bohm and D. Pines, Phys. Rev. 82, 625 (1951).
[25]D. Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1964).
[26]E. M. Conwell, Solid State Phys., Suppl. 9 (1967).
[27]W. P. Dumke, Phys. Rev. 118, 938 (1960).
[28]N. O. Folland, Phys. Rev. B 1, 1648 (1970).
[29]M. H. Jorgenson, Phys. Rev. B 18, 5657 (1978).
[30]J. P. Walter and M. L. Cohen, Phys. Rev. B 5, 3101 (1972).
[31]W. P. Dumke (unpublished).
[32]P. J. Price, Philos. Mag. 46, 1252 (1955).
[33]H. R. Shanks, P. D. Maycock, P. H. Sidles, and G. C Danielson, Phys. Rev. 130, 1743 (1963).
[34]E. J. Yoffa, Appl. Phys. Lett. 36, 37 (1980).
[35]For a review, see C. Jacoboni, C. Canali, G. Ottaviani, and A. Alberigi Quaranta, Solid-State Electron. 20, 77 (1977).

Role of carrier diffusion in lattice heating during pulsed laser annealing

Ellen J. Yoffa
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

Appendix H

(Received 21 August 1979; accepted for publication 23 October 1979)

A calculation is presented which demonstrates that diffusion of the hot, dense carriers generated in pulsed laser annealing of Si can substantially reduce the rate at which energy is transferred to the semiconductor lattice near the surface. The extent of the region in which this energy transfer occurs is consequently increased.

PACS numbers: 82.20.Rp, 72.20.Jv, 72.90. + y

Although pulsed laser annealing has been investigated for over a decade as a method of recrystallizing amorphized Si films,[1] the details of energy transfer from the laser to the semiconductor lattice remain a subject of debate.[2,3] Previous calculations[4-9] have relied on the important assumption that the laser energy is delivered to the lattice in the same region in which it is absorbed by the carriers, and that the resulting heat subsequently diffuses by conventional lattice and electronic thermal processes. In this letter we demonstrate that under conditions relevant to the laser annealing problem, excited carriers may diffuse an appreciable distance from the surface before they give their energy to the lattice.

The rate of change of the total carrier energy E is given by

$$\frac{\partial E}{\partial t} = g\hbar\omega_L \exp(-x/\delta) - \frac{N_e \hbar\omega}{\tau_e} + \frac{D_a E}{N}\frac{\partial^2 N}{\partial x^2}. \quad (1)$$

The first term represents the rate of laser energy input, where g is the photon absorption rate and δ is the absorption length at laser frequency ω_L. The second term takes into account the rate at which energy is given to the lattice by electrons having density N_e via emission of phonons with energy $\hbar\omega$. Phonon emission by holes, which involves intravalley transitions only, is effectively screened at high carrier densities. Although in general the emission frequency $1/\tau_e$ depends on the relative carrier and lattice temperatures and on the carrier density, for those densities pertinent to typical pulsed annealing experiments, τ_e is only a weak function of these quantities.[10] [In addition, because Eq. (1) will be solved for the steady-state condition of time-independent $E(x)$, the detailed dependence of τ_e on the total carrier energy is of minor importance.] We assume $\tau_e \sim 10^{-13}$ sec, which is consistent with luminescence data for crystalline Si analyzed by Folland[11] and with those values inferred from transport measurements.[12] Although the particular value of τ_e will vary under extreme conditions of implantation damage and/or amorphization, the weakly N-dependent nature of τ_e will most likely be preserved.

The final term in Eq. (1) represents diffusion of energy by actual carrier diffusion. Although energy is a conserved quantity, the number of carriers is not. When a particular carrier diffuses, it takes its energy with it, thereby reducing the total energy left behind and consequently lowering the carrier temperature. The carrier density therefore decreases not only as a direct result of carrier diffusion but also as an indirect result of the accompanying energy diffusion. Because the carrier density N varies strongly with carrier temperature T_e, $T_e \partial N/\partial x \gg N \partial T_e/\partial x$. We have therefore neglected diffusion terms involving $\partial D_a/\partial x$, where the ambipolar carrier diffusivity $D_a \simeq 2kT_e\tau_e\tau_h/(m_e^*\tau_h + m_h^*\tau_e)$, with $m_e^*(m_h^*)$ the electron (hole) effective masses, and $\tau_e(\tau_h)$ the carrier-phonon scattering times. The carriers diffuse at a rate $\partial N/\partial t \simeq D_a \partial^2 N/\partial x^2$; the average energy per carrier is E/N.

Although the carrier distribution in the steady state ($\partial E/\partial t = 0$) is time independent, the lattice temperature continuously increases as a consequence of the nonzero phonon emission term. "Steady state" therefore refers to the carrier system only. Using the energy conservation equation

$$\int_0^\infty g\hbar\omega_L \exp(-x/\delta)\,dx = \int_0^\infty \frac{N_{ss}(x)\hbar\omega}{\tau_e}\,dx \quad (2)$$

and the fact the E/N and τ_e are weak functions of N_e, we find the steady-state electron density

$$N_{ss}(x) \simeq \frac{g\hbar\omega_L \tau_e}{\hbar\omega}\,\frac{\delta/\alpha}{(1-\delta^2/\alpha^2)} \times\left[\exp(-x/\alpha) - \frac{\delta}{\alpha}\exp(-x/\delta)\right], \quad (3)$$

where $\alpha \equiv (D_a \tau_e E/N\hbar\omega)^{1/2}$. The effective diffusion length α has the form of a conventional diffusion length $(Dt)^{1/2}$, where here t (instead of being a recombination time) is the time during which a hot carrier can diffuse before it has given up its energy by phonon emission. Such a process requires $(E/N)(\hbar\omega)^{-1}$ collisions, so that $t = [(E/N) \times(\hbar\omega)^{-1}]\tau_e$. The fact that this result has such a simple interpretation suggests that it may be applicable in general provided an appropriate diffusion coefficient can be calculated.

The rate of energy loss by phonon emission can then be directly related to the rate of energy input by the laser. At the surface,

$$\frac{N_{ss}^0 \hbar\omega}{\tau_e} = \frac{g\hbar\omega_L}{(\alpha/\delta)+1}, \quad (4)$$

where $N_{ss}^0 \equiv N_{ss}(x=0)$. Thus we find that the rate of lattice heating at the surface is reduced from the laser energy input rate by a factor of $(\alpha/\delta + 1)$.

Numerical evaluation of Eq. (4) is complicated by the fact that α depends on the interrelated quantities N, T_e, and E. Collisions between carriers at densities exceeding $10^{[illegible]}$ cm ensure that electrons and holes rapidly establish a mutual temperature T_e which is initially much greater than the

 0003-6951/80/010037-02$00.50

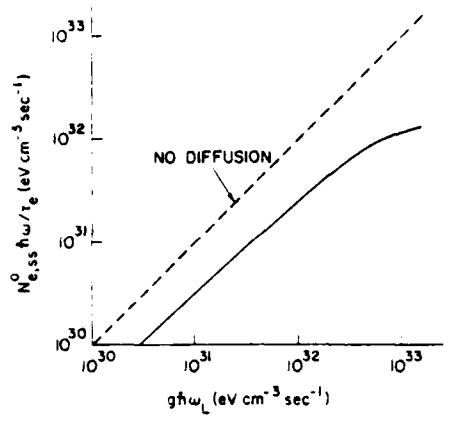


FIG. 1. Rate of phonon emission at the surface, $N^0_{e,ss}\hbar\omega/\tau_e$, as a function of the rate of laser energy input, $g\hbar\omega_L$. The dashed line was derived by neglecting carrier diffusion.

lattice temperature. In addition, the dominant recombination mechanism is Auger, which greatly reduces the generation-induced splitting of the electron and hole quasi-Fermi levels. Under these conditions, T_e and E are easily calculated for any N_e.

In Fig. 1 we plot Eq. (4) for $\hbar\omega_L = 2.3$ eV ($\delta \simeq 10^{-5}$ cm). When diffusion is neglected (indicated by the dashed line) the rate at which energy is transferred to the lattice at the surface equals the rate at which it is absorbed by the carriers. Under this condition, lattice heating occurs in a depth δ. However, we see from Fig. 1 that diffusion substantially reduces the rate at which energy is delivered to the lattice near the surface. When diffusion is important, lattice heating occurs in a depth which depends on both α and δ and is comparable to or larger than δ. For a typical annealing laser pulse[13] of 10^8 W/cm^2, $g\hbar\omega_L \simeq 3\times10^{31}$ eV/cm^3 sec, so that the steady-state electron density $N^0_{e,ss} = 5\times10^{19}$ cm^{-3}, which corresponds to $kT_e = 0.16$ eV and $E = 8\times10^{19}$ eV cm^{-3}. D_a is therefore $\simeq 100$ cm^2/sec, so that $\alpha \simeq 2.5\times10^{-5}$ cm and the lattice heating rate is reduced by a factor of approximately 3.5.

In summary, at high excitation rates, the laser energy is given to the lattice within a characteristic depth determined primarily by carrier diffusion. The effective diffusion length α, which is a function of the total carrier energy, increases monotonically with photon absorption rate, so that increasing g eventually leads to lattice heating over a larger region. In addition, a reduction in δ (e.g., by an increase in laser frequency) will not necessarily reduce the size of the heated volume, since for sufficiently small δ, carrier diffusion will eventually limit the minimum depth of heating. Because the ultimate temperature rise of the lattice depends so strongly on the extent of the region in which the energy transfer takes place, carrier diffusion plays a significant role in determining this final temperature.

This work was supported in part by the Air Force Office of Scientific Research under Contract No. F49620-77-C-0005. The author wishes to acknowledge J.A. Van Vechten for recognizing the importance of carrier diffusion and is grateful to R.C. Frye and P.J. Price for helpful discussions.

[1]H.R. Leuchtag, Physics Today **31** (7), (1978), and references therein.
[2]I.B. Khaibullin, E.I. Shtyrkov, M.M. Zaripov, R.M. Bayazitov, and M.F. Galjautdinov, Radiat. Eff. **36**, 225 (1978).
[3]J.A. Van Vechten, R. Tsu, F.W. Saris, and D. Hoonhout, Phys. Lett. (to be published).
[4]R.A. Ghez and R.A. Laff, J. Appl. Phys. **46**, 2103 (1975).
[5]J.C. Wang, R.F. Wood, and P.P. Pronko, Appl. Phys. Lett. **33**, 455 (1978).
[6]M. Lax, Appl. Phys. Lett. **33**, 786 (1978).
[7]J.C. Schultz and R.J. Collins, Appl. Phys. Lett. **34**, 84 (1979).
[8]A. Lietoila and J.F. Gibbons, Appl. Phys. Lett. **34**, 335 (1979).
[9]C.M. Surko, A.L. Simons, D.H. Auston, J.A. Golovchenko, R.E. Slusher, and T.N.C. Venkatesan, Appl. Phys. Lett. **34**, 635 (1979).
[10]A detailed calculation of the effects of a hot, dense plasma on the rate of phonon emission will be presented in: E.J. Yoffa, Phys. Rev. B (to be published).
[11]N.O. Folland, Phys. Rev. B **1**, 1648 (1970).
[12]For a review, see C. Jacoboni, C. Canali, G. Ottaviani, and A. Alberigi Quaranta, Solid State Electron. **20**, 77 (1977).
[13]R. Tsu, J.E. Baglin, T.Y. Tan, M.Y. Tsai, K.C. Park, and R. Hodgson, AIP Conf. Proc. No. 50 (1979), p. 344.

Contribution of Dipole Defects to DLTS Spectrum*

J. A. Van Vechten, C. M. Ransom, and T. I. Chappell

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598
United States of America

Appendix I

Deep Level Transient Spectra (DLTS) have previously been analyzed assuming that the only contributions to the variations in junction capacitance recorded come from detrapping of carriers singly from monopolar defects. The concentration of defects deduced in this way is too small to account for measured carrier lifetimes. It is also smaller than is deduced from theory and from TEM. We argue that many deep level defects are clusters of point defects whose electrical activity results in dipole, and higher multipole charge states. Transitions among these states can also contribute to the DLTS signal but much more weakly. Consequently, conventional analysis grossly underestimates the number of defects present.

I. Motivation

DLTS spectra have generally been analyzed [1] as if the deep level defects being observed were simple point defects, although it has been widely recognized that this is not true[2-4]. In samples that have seen room temperature or above, one may expect to find complexes of point defects formed from the supersaturations of defects introduced by high temperature processing, e.g., by crystal growth, or by irradiation. This is because simple point defects are generally ionized and rather mobile at room temperature and readily form coulombically or elastically bound clusters. Nevertheless, in DLTS analysis it has been assumed that the observed variation in junction capacitance comes only from the trapping and detrapping of single carriers at the various defect centers. Some justification for the simple form of analysis has come from the fact that experiments have been designed with the intent to trap only holes, h_v^+, or only electrons, e_c^-. Experimentalists have assumed that by limiting minority carrier densities, they succeeded in preventing possible contributions to the DLTS signal through recombination, reorientation, or de-excitation of dipolar and higher multipolar charge distributions at the defects.

There is much reason to believe that the conventional analysis of DLTS spectra has significantly underestimated the concentrations of deep defects present. Although the analysis supposedly gives defect concentrations, cross sections and activation energies, attempts to calculate carrier lifetimes from DLTS measurements generally overestimate these lifetimes rather badly when proper account is taken of the expected saturation of the traps [5,6]. A generally successful theory of the thermochemistry of semiconductor defects[4] predicts substantially higher concentrations of defect clusters than have been inferred from DLTS. For example, in compound semiconductors like GaP and GaAs, it is predicted that the cluster consisting of one antisite defect, e.g., a P on a Ga site, P_{Ga}, bounded by two vacancies, e.g., V_{Ga}, to form a (110) oriented complex, e.g., $V_{Ga}^- P_{Ga}^{+2} V_{Ga}^-$, one atom in diameter and 3 lattice sites long should typically occur in mid 10^{17} cm^3 concentrations even in the best quality material[7,8]. DLTS workers have not reported any spectra they could associate with such a defect at anywhere near that high a concentration. On the other hand, ultrahigh resolution, direct lattice imaging TEM does find defect clusters having this size, shape, orientation and concentration in the best quality GaAs and it has been shown that they are active in infrared absorption[9]. It can also be argued that such concentrations of defects are necessary to account for the degradation of electrooptic devices[8,10].

Thus, the motivation of this paper is to understand why DLTS studies have not reported as many deep level defects as have been predicted from theory or inferred from electrical measurements, TEM, and observations of degradation phenomena. In Sec. II we make introductory remarks about experiments, defects, and defect interactions. In Sec. III we discuss what processes the defect clusters that we should expect to be present might exhibit which would produce a DLTS signal much weaker than that produced by simple detrapping. In Sec. IV we explain why the experimental design failed to count all the defects that were there by detecting detrapping and failed to prevent the operation of the weaker

processes. Thus, one defect may produce more than one DLTS signature and the concentration deduced from each may be much too small.

II Distinctions among Experiments and among Defects; Interaction among Defect Ionization Levels

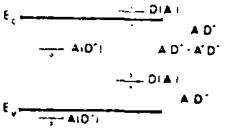
There are two kinds of donors and acceptors. There are those, which we call coulombic, which are neutral when filled by free carriers, like S_P and Zn_{Ga} in GaP, and there are those, which we call *isoelectronic*, which are ionized when filled by free carriers, like Bi_P and N_P in GaP. Vacancies and interstitials are important examples of isoelectronic defects. In recognition of this fact DLTS experimentalists report "h_V^+-traps" and "e_C^--traps". The defect clusters that form during cooling from crystal growth/processing temperatures will probably contain combinations of both types. The $V_{Ga}P_{Ga}V_{Ga}$ center is an example. Such combinations are likely because coulombic defects, like dopants, usually migrate via vacancy or interstitial mechanisms. Also the clusters can accumulate several point defects during extended periods of atomic migration. Thus, the larger clusters are apt to be able to both trap and emit carriers of both type in transitions between states which may or may not have the same net charge.

We distinguish two kinds of DLTS experiment: ACDLTS, in which the AC capacitance is measured[1], and DIDLTS, in which the total displacement current integrated through the rate window interval, RWI, is measured[11]. Suppose that an e_C^- h_V^+ pair is bound at a defect site giving it a dipole moment and contributing to its polarizability. When the recombination of such a pair is thermally activated, as is usual for non-radiative deep levels, it will produce an ACDLTS signal due to the change in the polarizability of the center in the junction field and a DIDLTS signal due to the change in dipole moment. However, the magnitude of this signal per defect will be weaker by a factor of order 10^{-4} than that which would occur if there were simple detrapping of a carrier. Using the expression for the dipole contribution to the polarization of the dielectric, $\Delta\epsilon/\epsilon = N\mu^2/3kT$ and the relation $\Delta C/C = \Delta\epsilon/2\epsilon$, it is possible to calculate that 10^{18} dipoles having unit electron charges separated by 0.3 nm will produce the same magnitude DIDLTS signal as only 10^{14} detrapping defects. The ACDLTS signal would depend upon the spring constants at the dipole. The concentration of defects ascribed to this DLTS signature would be underestimated by a factor of order 10^4; the relative magnitude of the signal will be different for ACDLTS and DIDLTS. If instead the pair dissociates releasing both carriers from the defect and both are swept entirely out of the depletion region, DIDLTS will see a full detrapping signal from pair emission, but ACDLTS will record only the weak signal because the net charge in the depletion region does not change. For the case when both carriers are captured at other defect sites before they can be swept out, DIDLTS will also record only a weak signal. If one carrier is swept out and one is retrapped, both methods will record a strong signature due to the one that got out and a weak signature due the subsequent behavior of the one retrapped.

Ionization levels of defects in close association are not independent of the proximity and ionization state of other nearby defects. We denote a two defect complex in some semiconductor host AD_ζ where A and D are point defects which singly would have acceptor or donor ionization levels, $^{-}_{0}A$ and $^{0}_{+}D$, in the gap respectively. (Of course, many defects, such as V_{Si}, have both donor and acceptor levels.) The proximity of a neutral donor, D^0, generally lowers $^{-}_{0}A$ and vice versa. The proximity of an ionized donor, D^+, greatly lowers $^{-}_{0}A$ and vice versa. We denote the ionization levels $^{-}_{0}A(D^0)$ and $^{-}_{0}A(D^+)$ for the acceptor and $^{0}_{+}D(A^0)$ and $^{0}_{+}D(A^-)$ in the accompanying figure. The magnitude of this spread for ionized point defects is $e^2/\epsilon R + L_\lambda(D^+)$, where ϵ is the effective dielectric constant, R the spacing, and $L_\lambda(D^+)$ is the lattice relaxation at the ionized acceptor in the presence of the ionized donor. For nearest neighbor pairs the coulombic, first term alone would be about 0.5 eV. Lattice relaxation contributions may sometimes be even larger[12].

If the isolated defects are shallow, e.g., Zn - S pairs in GaP [13], close association drives the ionization levels into the bands, i.e., $^{0}_{+}D(A^-)$ touches the conduction band edge, E_c, and $^{-}_{0}A(D^+)$ touches the valence band edge, E_v. Then the neutral states of A and D become metastable. For any position of E_f, the equilibrium state of the center, A^-D^+ has a dipole, but no monopole, moment. It would never be seen detrapping in DLTS although it would scatter carriers. If the ionization levels of the isolated point defects are deeper, then the spread of individual ionization levels, i.e., $^{-}_{0}A(D^0) - ^{-}_{0}A(D^+)$ and $^{0}_{+}D(A^-) - ^{0}_{+}D(A^0)$ may not touch either E_c or E_v. For Cd - O close pairs in GaP the acceptor spread touches E_v but the donor spread does not touch E_c [14]. Thus, the complex can stablely bind an e_C^- and maintain a negative state, A^-D^0, when E_f is sufficiently high. Otherwise, it is neutral with a dipole moment, A^0D^+. It will scatter carriers and have some activity as a recombination center. It would be seen detrapping in DLTS for n-type material, but not for p-type material.

Suppose now that $^{-}_{0}A(D^{0})$ lies above $^{+}_{0}D(A^{0})$ as is common for pairs of isoelectronic point defect (It is true for V_{Si}.) Because the spread between $^{-}_{0}A(D^{0})$ and $^{-}_{0}A(D^{+})$ and between $^{0}_{+}D(A^{-})$ and $^{0}_{+}D(A^{0})$ is so large for close pairs, greater than 0.5 eV, they may overlap and together span a large fraction of the band gap. When this occurs, then there is a large range of E_f over which the clusters are stable either as $A^{0}D^{0}$ or as $A^{-}D^{+}$, but not as $A^{-}D^{0}$ or as $A^{0}D^{+}$. The relative proportion of these two neutral states (for larger clusters there would be several) depends upon E_f as well degeneracy factors. In the accompanying figure we show the case that the overlapping ionization levels touch both band edges. Such a defect could only be stable when neutral, binds carriers only in pairs, and may or may not have a dipole moment. It could be a very active recombination center.



III Alternative Mechanisms Changing Junction Capacitance

We discussed the recombination of e_c^- h_v^+ pairs above. Deep level defects are particularly important precisely because many of these act as recombination centers. High recombination activity suggests that both e_c^- and h_v^+ are trapped at the defect so that their wave functions overlap strongly. This in itself implies that the relevant defects have a dipole rather than a monopolar charge distribution. When it is thermally activated, internal rearrangement of charge within a cluster will produce a DLTS signature The three neutral states of $V_{Ga}P_{Ga}V_{Ga}$, $V^{-}_{Ga}P^{+2}_{Ga}V^{-}_{Ga}$, $V^{0}_{Ga}P^{+}_{Ga}V^{-}_{Ga}$, and $V^{0}_{Ga}P^{0}_{Ga}V^{0}_{Ga}$, each have different polarizabilities and ranges of E_f for which they will be stable. Note the the first and last have no dipole moment while the middle does and that all bind 2 e_c^-'s; no trapping or detrapping of free carriers is involved in these transitions. However, any of the neutral states would be an effective nonradiative recombination center for h_v^+'s.

Many deep levels exhibit recombination enhanced diffusion[15], RED. Recombination may alter either the magnitude or the orientation of a cluster's dipole moment. Even if it does not alter the field free dipole moment, atomic reconfiguration of the cluster will likely change its polarizability. The pair recombination event in RED is apt to leave the cluster in a distorted, nonequilibrium configuration. Then thermally activated atomic migration would be expected to return the defect to its equilibrium configuration. Any such reconfiguration will change its polarizability and produce an ACDLTS signal. Reorientation of a dipole would also produce an additional contribution to the DIDLTS signal unless the dipole moments were randomly oriented and their rate of reconfiguration were independent of orientation.

Now we consider a mechanism which requires only majority carriers and a change in charge state of at least one member of the defect cluster, as would be caused by the flow of current or the shift in E_f between the trap filling and measurement phases of the DLTS experiment. As many deep defects have different equilibrium configurations in different charge states, a change in charge state can cause atomic motion. When the original charge state is regained, atoms may go to different but formally equivalent lattice locations. This is known as the Bourgoin-Corbett, B-C, mechanism and produces rapid, athermal migration[16]. This, too, may reconfigure the defect cluster in a way that changes its dipole moment. As it regains its equilibrium configuration by thermally activated atomic motion during the measurement phase, a DLTS signal is produced.

If the junction field is strong enough at the measurement temperature to align defects which normally have a dipole moment, then thermal migration caused either by the B-C mechanism or by RED during the trap filling stage would randomize their orientation. One would see a DLTS signal due to the thermally activated motion of reorientation. As the junction fields are of order 10^{4} eV/cm and dipole lengths are of order 10^{-7} cm, the relative fraction of such centers aligned would normally be small, e.g., 4 % at room temperature. Thus, the relative magnitude of this contribution to that which would have resulted from thermally activated detrapping is even smaller by that factor. The underestimate of the number of such centers present would be correspondingly greater.

IV How might Monopolar Contributions be Missed; How Limited is Minority Carrier Concentration?

The existence of the mechanisms just described is not in doubt. On the other hand, in order to show that they may have produced a significant error in the estimate of the number of deep level defects present, we must show why an experiment designed to affect and detect the detrapping of carriers singly from the expected defect clusters should have failed to do so. If the monopolar effect were seen, it would be so much larger than the dipolar that we would merely have reported a few too many defects.

Of course, we bear in mind that any athermal process, such as recombination by tunneling, will go undetected in DLTS.

As was noted above, if the level scheme indicated in the figure occurs, then the center will only bind carriers in pairs, pulling an electron out of the valence band if necessary to fill the $^{-}_{0}A(D^{+})$ level after D has attained the D^{+} by either emitting an e^{-}_{c} or trapping a h^{+}_{v}. Changes in net charge state would always be over before the RWI opens and the center would be undercounted by the factor 10^{-4}.

For B-C migration we would need to suppose that the center has returned to its original charge state before the rate window opens during the measurement phase. This would occur only if the carrier is trapped in a state that is metastable and therefore likely to be occupied only when current is flowing during the trap filling and transient stages or if the filled state is so shallow that it empties so soon after E_f passes through it that the detrapping is not recorded. If Cd exhibited B-C migration in GaP, the Cd-O pair in n-type GaP would be an example of this. Point defects known to exhibit B-C migration are rather deep[16], but in forming clusters their levels could spread to the band edge.

In addition to the minority carriers running free in their band, which the experimentalist controls with bias voltage (or optical pumping), one must also consider those bound in the defect complexes. When raising E_f to inject e^{-}_{c}'s to be trapped, one will also release some concentration of h^{+}_{v}'s, that may be comparable with the total number of point defects present. These can bind at the same sites which just trapped the e^{-}_{c}'s to neutralize the center again. Recombination and all the effects associated with it above may follow. Another important point is that some defects, in particular the single vacancy[17] in Si, V_{Si}, exhibit a "negative U" for certain ionization states. That is V^{0}_{Si} and V^{+2}_{Si} are stable, but V^{+}_{Si} is not stable. This means that the trapping of a hole at a V^{0}_{Si} will be followed by an athermal process in which an e^{-}_{c} is emitted and V^{+2}_{Si} results. Thus, one minority carrier is generated for each majority carrier trapped in such "negative U" defects. These effects would cause under counting by some statistical factor less severe than 10^{-4}.

Impact ionization at defect centers by carriers accelerated by the transient fields of the experiment is possible. Since fields as high as 10^{4} V/cm exist over regions as wide as 1 μm, carriers may be excited to energies reaching 1 eV. During injection a fast majority carrier might knock a minority carrier free. It also might knock a majority carrier out of a deep trap that would normally be filled during the measurement phase. This could shift the level of a minority trap to its band edge. Thus, impact ionization could provide minority carriers to be trapped in the region under test at concentrations a significant fraction of the total number of carriers swept through, which would be comparable with the background doping level.

V References

* Supported in part by the Air Force Office of Scientific Research contract No. F49620-79-C-0077.

1) D. V. Lang, J. Appl. Phys. 45 (1974) 3023.
2) P. M. Mooney, L. J. Cheng, M. Suli, J. D. Gerson, and J. W. Corbett, Phys. Rev. B 15 (1977) 3836.
3) Y. H. Lee, K. L. Wang, A. Jaworowski, P. M. Mooney, L. J. Cheng and J. W. Corbett, Phys. Stat. Solidi (a) 57 (1980) 291.
4) J. A. Van Vechten in "Handbook of Semiconductors Vol. 3 Materials Properties and Preparation", edited by S. P. Keller (North Holland, Amsterdam, 1980) Chapter 1.
5) D. V. Lang and R. A. Logan, J. Electron. Mater. 4 (1975) 1053 and private communication.
6) B. Hamilton, A. R. Peaker, and D. R. Wight, J. Appl. Phys. 50 (1979) 6373.
7) J. A. Van Vechten, J. Electron. Mat. 4 (1975) 1159.
8) J. A. Van Vechten, J. Electrochem. Soc. 122 (1975) 1556.
9) J. B. Van der Sande and E. T. Peters, J. Appl. Phys., 45 (1974) 1298.
10) P. Petroff, W. D. Johnston and R. L. Hartman, Appl. Phys. Lett. 25 (1974) 226.
11) J. A. Barsuk and R. M. Swanson, IEEE Trans. Electron Devices, Dec. 1980.
12) G. D. Watkins, Inst. Phys. Conf. Ser. 23 (1975) 1.
13) P. J. Dean, C. J. Frosh and C. H. Henry, J. Appl. Phys. 39 (1968) 5631.
14) T. N. Morgan, B. Welber and R. N. Bhargava, Phys. Rev. 166 (1968) 751.
15) L. C. Kimerling, Inst. Phys. Conf. Ser. 46 (1979) 56.
16) J. C. Bourgoin and J. W. Corbett, Phys. Lett. 38A (1972) 135.
17) G. D. Watkins and J. R. Troxell, Phys. Rev. Lett. 44 (1980) 593.

PHYSICAL REVIEW B VOLUME 21, NUMBER 8 15 APRIL 1980

Scattering-theoretic method for defects in semiconductors. II. Self-consistent formulation and application to the vacancy in silicon

J. Bernholc, Nunzio O. Lipari, and Sokrates T. Pantelides
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 15 November 1979)

A self-consistent-field method for calculation of the electronic structure of localized defects in semiconductors is described. The method is based on Green's-function theory and follows the original idea of Koster and Slater and its developments by Callaway and coworkers. The Wannier functions of the original formulations are, however, replaced by a more flexible set of linear combination of atomic orbitals. This choice and an accurate evaluation of the perfect-crystal Green's function bring this method to the level of sophistication, accuracy, and rigor characteristic of state-of-the-art band-structure and surface calculations. The efficiency of the method stems largely from the fact that it exploits both the translational symmetry of the host crystal and the short range of the defect potential. Thus, all bulk properties (e.g., band gaps, bandwidths, etc.) are built in from the start via a band-structure calculation and are preserved. One then focuses on the changes produced by the defect potential, so that the interpretation of the results is straightforward and unambiguous. In this paper, we report an application of this method to an isolated vacancy in Si assuming no lattice relaxation. The unrelaxed vacancy introduces a bound state of T_2 symmetry at 0.7 eV above the valence-band edge and a number of resonances and antiresonances within the valence bands. A detailed analysis of these states in terms of their origin, orbital content, and of state and charge densities is presented. We find that, while many of these states are individually quite extended, they combine destructively to produce a very localized net change in the charge density. We also find that the resulting localized potential can be well approximated by a negative of an atomic silicon potential extracted from a self-consistent bulk calculation. Finally, we compare the relative merits of the three increasingly more sophisticated, but also more costly, approaches to the defect problem, namely, (1) tight-binding, (2) non-self-consistent, and (3) self-consistent calculations.

I. INTRODUCTION

The calculation of the electronic structure and properties of a perfect crystalline semiconductor is at present routinely carried out by using band-theoretic techniques, which exploit translational periodicity (Bloch theorem). Defects, on the other hand, break the translational symmetry of the host crystal, and, as a result, the calculation of their electronic properties is a considerably more complicated problem.

One class of point defects is treated very successfully by the well-known effective-mass theory (EMT).[1] The theory works best for shallow donors and acceptors whose perturbation potential is dominated by the screened Coulombic tail responsible for the hydrogenic spectra near band edges. The corresponding wave functions are very extended in real space and hence highly localized in $\vec{k}$ space. Recently, new developments in EMT suggest that some moderately deep levels can also be handled by similar techniques.[2-4] For many deep levels, however, the EMT assumptions are not suitable and alternative techniques are necessary to describe wave functions which are highly localized in real space.

Two distinct approaches have thus far been pursued for the study of deep levels in semiconductors[5]:

One of them approximates a defect in a perfect infinite solid with a finite cluster. The Schrödinger equation for the cluster is then solved directly. A variety of techniques differing in the way the cluster is terminated have evolved: free clusters,[6-8] saturated clusters,[9,10] repeated clusters,[11,12] etc. Most of the work has been carried out on clusters consisting of 4–70 atoms. Clusters of that size, however, do not contain the usually quite extended defect wave function[7,11,12] and are therefore only suitable for qualitative analysis. Recently, techniques became available for calculating local densities of states of clusters of 2000 or more atoms.[13,14] These techniques, however, have thus far been used only with semiempirical tight-binding Hamiltonians. Furthermore, none of the cluster methods has proven to be particularly efficient. The main shortcoming of all these methods is that they must produce not only the defect-induced states, but also all the bulk states of the host, without exploiting the underlying translational periodicity.

In the second approach, the main shortcoming of the cluster methods is eliminated by first calculating the host-crystal properties using band theory and then focusing on the defect-induced changes, i.e., bound states, resonances, antiresonances, etc. A variety of techniques have evolved along these lines as well.[1,15-17] With the

exception of effective-mass theory, all other techniques are variants of a method that is usually attributed to Koster and Slater.[17] These techniques are in principle especially efficient when the perturbation potential U is *localized*, even though the individual wave functions may be extended. They therefore complement EMT-type theories, which rely on U having a dominant Coulombic tail.

Since the time of the original papers, the Koster-Slater method has been generalized in the language of scattering theory and Green's functions by Callaway.[18] Both Koster and Slater[17] and Callaway[18] developed the theory by introducing the host-crystal Wannier functions to expand wave functions and represent operators in matrix form. Callaway and Hughes[19] subsequently applied the method to the single vacancy in Si, but the construction of the Wannier functions proved to be very tedious and difficult, thus severely limiting the efficiency of the method. For that reason only a few other applications[20-22] have been made. A simplified tight-binding description of the vacancy in diamond and silicon was subsequently reported by Lannoo and Lenglart[23] and by Rouhani *et al.*,[24] but the approach was not pursued further.

More recently, a variant of the Koster-Slater method introduced by Bassani, Iadonisi, and Preziosi (BIP) (Ref. 25) has been used extensively by Jaros and Brand[26] and by Lindefelt.[27] The relationship between the BIP method and the Koster-Slater method is discussed in Refs. 5 and 28. Even more recently, there has been renewed interest in the original Koster-Slater method. Bernholc and Pantelides[28] (paper I) discussed the use of an operator formulation which reveals clearly that the role of the Wannier functions is simply to represent operators in matrix form. Thus, instead of the Wannier functions, other more convenient sets can be used. A particularly convenient choice is a set of linear combinations of atomic orbitals (LCAO) which are appropriate to carry out a band-structure calculation for the host material. Using the same LCAO set for the point-defect problem is then equivalent to expanding the perturbed-crystal wave functions in terms of the host-crystal Bloch functions or Wannier functions. In paper I,[28] empirical tight-binding Hamiltonians were used to study the ideal vacancy (simple removal of an atom without lattice reconstruction) in Si, Ge, GaAs. Since then, we have extended the same basic idea to include accurate self-consistent pseudopotential Hamiltonians for the host and to allow for electronic-charge redistribution that accompanies the removal of an atom, i.e., self-consistency for the defect as well. The resulting self-consistent Green's-function method for point defects has been used to obtain a detailed description of the unrelaxed vacancy in Si. The main results have been reported in a Letter.[29] The publication of that Letter coincided with the publication of a Letter by Baraff and Schlüter[30] who reported the development of a similar LCAO self-consistent Green's-function method and an application to the unrelaxed vacancy in Si.

In this paper we give a full account of our formulation of the self-consistent Green's-function method and its application to the unrelaxed vacancy in Si. This work brings the method to the same level of sophistication, accuracy, and resolving power as achieved by the most recent band-structure and surface calculations. In particular, the iteration to self-consistency frees our results from any dependence on the assumed similarity of interactions in the perturbed system to those in the unperturbed bulk crystal.[29] In this work we prove for the first time that the vacancy potential is very *localized*, while the individual vacancy states are quite *extended*. We also provide charge-density maps of the various states and make a detailed analysis of the electron distribution in the vicinity of the vacancy. Finally, we examine the effects of self-consistency and propose the bulk silicon potential to be used in inexpensive semiquantitative studies of vacancy complexes in silicon.

This paper is organized as follows. In Sec. II, the general Green's-function formalism is described. This section is based mainly on the papers by Koster and Slater,[17] Callaway,[18] and on general results of formal scattering theory.[31,32] The representation of operators is dealt with in Sec. III and the choice of basis functions is discussed in Sec. IV. Our computational procedures are outlined in Sec. V. Section VI contains the results of the unreconstructed vacancy in silicon and is followed by a brief summary and conclusions (Sec. VII). In a series of appendices we give additional details of our calculations and compare our formulation and results with those of Ref. 30.

II. SELF-CONSISTENT GREEN'S-FUNCTION METHOD

The Green's-function method assumes the knowledge of the solutions of the perfect-crystal Hamiltonian H^0:

$$H^0\psi^0_{n\vec{k}} = E_{n\vec{k}}\psi^0_{n\vec{k}} , \tag{1}$$

where $E_{n\vec{k}}$ and $\psi^0_{n\vec{k}}$ denote the band energies and wave functions, respectively. The corresponding Schrödinger equation for the imperfect solid is

$$H\psi = (H^0 + U)\psi = E\psi . \tag{2}$$

We define the perfect- and perturbed-crystal Green's operators by the relations

$$G^0(E^+) = \lim_{\epsilon \to 0^+} (E + i\epsilon - H^0)^{-1} \tag{3}$$

and

$$G(E^+) = \lim_{\epsilon \to 0^+} (E + i\epsilon - H)^{-1} . \tag{4}$$

Combining Eqs. (2)–(4), one immediately gets Dyson's equation, i.e.,

$$G(E^+) = G^0(E^+) + G^0(E^+)UG(E^+) . \tag{5}$$

The formal solution of Eq. (5) is

$$G(E^+) = [1 - G^0(E^+)U]^{-1} G^0(E^+) . \tag{6}$$

Using the above definitions, one can immediately obtain formal solutions of Eq. (2). For energies within the band gaps, where $G^0(E)$ is real, the solution is

$$\psi = G^0(E)U\psi . \tag{7}$$

Within the band continua, one must add a solution of the homogeneous Eq. (1) to Eq. (7). We thus obtain the Lippman-Schwinger equation[33]

$$\psi_{n\vec{k}} = \psi^0_{n\vec{k}} + G^0(E^+)U\psi_{n\vec{k}} \tag{8}$$

(n and $\vec{k}$ are not conserved in the perturbed crystal, but are convenient labels for the scattering states). The condition for the existence of bound states is, from Eq. (7)

$$D(E) = \det\|1 - G^0(E)U\| = 0 , \tag{9}$$

when the operator $1 - G^0(E)U$ is expanded in any complete set of states. This condition determines the energy of the bound state. The corresponding wave function is then obtained as the nontrivial solution to Eq. (7) at that energy. For states within the band continua, we write Eq. (8) in the form

$$[1 - G^0(E^+)U]\psi_{n\vec{k}} = \psi^0_{n\vec{k}} . \tag{10}$$

The determinant of $1 - G^0(E^+)U$ is now nonzero because the imaginary part of $G^0(E^+)$ is nonzero within the band continua. As a result, the operator is always invertible and solutions exist at all energies within the energy bands of the perfect crystal. Band edges, therefore, are not shifted by the perturbation. The wave function ψ corresponding to a solution at $E = E^0_{n\vec{k}}$ is not, of course, equal to $\psi^0_{n\vec{k}}$.

The density operators for the perfect and perturbed crystal are related to the respective Green's operators by[31,32]

$$\rho^0(E) = -(2\pi)^{-1} \operatorname{Im} G^0(E) \tag{11}$$

and

$$\rho(E) = -(2\pi)^{-1} \operatorname{Im} G(E) . \tag{12}$$

Using Eq. (6), we now obtain the following expression[31] for the change in the charge density $\Delta\rho$ induced by the defect

$$\Delta\rho = \int_{-\infty}^{E_F} [\rho(E) - \rho^0(E)]\, dE$$
$$= \frac{2}{\pi} \operatorname{Im} \int_{-\infty}^{E_F} \{1 - [1 - G^0(E)U]^{-1}\} G^0(E)\, dE . \tag{13}$$

This expression is of central importance for self-consistent calculations. Similarly, the state density $N(E)$ is altered in the vicinity of the defect. Using the relations[31]

$$\Delta N(E) = \operatorname{Tr}[\rho(E) - \rho^0(E)] , \tag{14}$$

$$[G^0(E)]^2 = -(d/dE) G^0(E) , \tag{15}$$

and Eqs. (6), (11), and (12), one obtains

$$\Delta N(E) = (2\pi)^{-1} \operatorname{Im} \operatorname{Tr}\{(d/dE)[G^0(E)]\, U[1 - G^0(E)U]^{-1}\} \tag{16}$$

by cyclic properties of the trace. Since the derivative of the density of states is infinite at a critical point, one should in general expect some structure in $\Delta N(E)$ at those points (dependent on the strength and details of the potential). The potential will also introduce *additional* structure through the factor $[1 - G^0(E)U]^{-1}$. It can be shown[13] that the change in the state density is also given by

$$\Delta N(E) = \frac{2}{\pi} \frac{d\delta(E)}{dE} , \tag{17}$$

(spin included) where the phase shift $\delta(E)$ is defined by

$$\delta(E) = -\tan^{-1}[\operatorname{Im} D(E) / \operatorname{Re} D(E)] . \tag{18}$$

It follows that $\delta(E)$ goes through an odd multiple of $\pi/2$ every time $\operatorname{Re} D(E) = 0$. An expansion around such an energy E_0 gives

$$\tan\delta(E) \approx -\Gamma/[2(E - E_0)] , \tag{19}$$

where

$$\Gamma = 2 \operatorname{Im} D(E_0) / \operatorname{Re} D'(E_0) \tag{20}$$

and the prime denotes differentiation with respect to energy. In the region close to E_0, $\Delta N(E)$ becomes

$$\Delta N(E) = \frac{\Gamma}{2\pi} \frac{1}{(E - E_0)^2 + \frac{1}{4}\Gamma^2} . \tag{21}$$

This characteristic Breit-Wigner form indicates that for $\Gamma > 0$ one has a *resonance* or a peak in $\Delta N(E)$ with a half width Γ. For $\Gamma < 0$, one has an *antiresonance*, i.e., a negative peak with half width $|\Gamma|$. (This analysis applies only for isolated resonances or antiresonances. If the spacing between a resonance and an antiresonance is smaller than the width, a more complicated spectrum oc-

curs.) The above considerations also assume that the background state density (i.e., the state density of the perfect crystal) is smooth. The resonances and antiresonances associated with the critical points, i.e., originating from the factor $(d/dE)G^0(E)$ in (16), will be called quasiresonances and quasiantiresonances, respectively.

According to Levinson's theorem[31] the total number of states remains unchanged in the presence of a perturbation, i.e.,

$$\int_{-\infty}^{\infty} \Delta N(E)\,dE = 0 . \tag{22}$$

When the states in the gaps are counted separately from the state density changes within the bands, Eq. (22) becomes

$$\int_{\text{bands}} \Delta N(E)\,dE = -N_b , \tag{23}$$

where N_b is the total number of bound states in the gaps.

Let us consider a continuous group of bands separated by band gaps, e.g., the valence bands of a semiconductor. Denoting by E_{bot} and E_{top} the bottom and the top of this group of bands, one obtains from (17)

$$\int_{E_{\text{bot}}}^{E_{\text{top}}} \Delta N(E)\,dE = \frac{2}{\pi}\left[\delta(E_{\text{top}}) - \delta(E_{\text{bot}})\right] . \tag{24}$$

Since the Green's function and therefore also the determinant $D(E)$ are real at the band edges it follows from (18) that

$$\int_{E_{\text{bot}}}^{E_{\text{top}}} \Delta N(E)\,dE = 2m , \tag{25}$$

where m is in integer, (including spin) independently of either the strength or the details of the potential.

III. REPRESENTATION OF OPERATORS

For applications to particular problems, the operators of the preceeding section must be represented in a basis set. The operator equations (7), (9), and (13) then become matrix equations. In order to exploit the limited range of the defect potential, we will assume that the individual basis functions are localized in space (for example, Wannier functions or LCAO's centered at the atomic positions). The subject of the specific choices of the basis sets will be discussed in the next section. In such a representation $\{\varphi_\alpha\}$, the matrix elements of the potential $U_{\alpha\beta}$ will be nonzero only if both φ_α and φ_β are centered on atoms close to the defect, so that they both overlap with the potential. The space spanned by this basis set can therefore be divided into two subspaces: subspace A which overlaps with the potential and subspace B in which the potential is effectively zero. The potential matrix may then be written schematically as

$$U = \begin{pmatrix} U_{AA} & 0 \\ 0 & 0 \end{pmatrix} . \tag{26}$$

(Note that the size of the subspace A depends also on the range of energies one wishes to study. For high-energy scattering, a larger subset would be necessary.) After writing the Green's-function matrix as

$$G^0(E) = \begin{pmatrix} G^0_{AA}(E) & G^0_{AB}(E) \\ G^0_{BA}(E) & G^0_{BB}(E) \end{pmatrix} , \tag{27}$$

the matrix of the operator $1 - G^0(E)U$ becomes

$$1 - G^0(E)U = \begin{pmatrix} 1 - G^0_{AA}(E)U_{AA} & 0 \\ -G^0_{BA}(E)U_{AA} & 1 \end{pmatrix} . \tag{28}$$

From (28) it is now evident that

$$\det\|1 - G^0(E)U\| = \det\|1 - G^0_{AA}(E)U_{AA}\| , \tag{29}$$

so that the size of the determinant reduces to the size of the nonzero part of the potential matrix.[17] Let us write the bound-state wave functions as $\{\psi_A, \psi_B\}$, where ψ_A and ψ_B are the components of ψ in the subspace A and subspace B, respectively. If E_0 is the bound-state energy, ψ_A is the nontrivial solution to the matrix equation

$$\left[1 - G^0_{AA}(E_0)U_{AA}\right]\psi_A = 0 ; \tag{30}$$

ψ_B can then be obtained from [cf. Eq. (7)]

$$\psi_B = G_{BA}(E_0)U_{AA}\psi_A . \tag{31}$$

The correct normalization of ψ_A can be obtained without calculating ψ_B (Refs. 34, 30) as follows: From Eq. (7) we get

$$1 = \langle\psi|\psi\rangle = \langle\psi_A U G^0(E)|G^0(E)U|\psi\rangle$$
$$= \langle\psi_A U\left[G^0(E)\right]^2 U|\psi\rangle . \tag{32}$$

Using the relation

$$U|\psi\rangle = \{U_{AA}\psi_A, 0\} \tag{33}$$

and Eqs. (15) and (27), one obtains the normalization condition

$$-\langle\psi_A|U_{AA}G^0_{AA}(E)U_{AA}|\psi_A\rangle = 1 . \tag{34}$$

Since the self-consistent total change in the charge density cannot extend further than the defect potential,[35] it is sufficient to calculate $\Delta\rho$ in Eq. (13) only in the subspace A. It follows from Eqs. (26), (27), and (28)

$$\Delta\rho^{\text{val}}_{AA} = \text{Im}\,\frac{2}{\pi}\int_{\text{val bands}} dE\{1-[1-G^0_{AA}(E)U_{AA}]^{-1}\} \times G^0_{AA}(E). \quad (35)$$

The relations (29), (30), (34), and (35) show explicitly that the size of the problem in the Green's-function method is determined by the range of the perturbation potential even for self-consistent calculations.

In order to complete the operator formalism for the present Green's-function method, we give some general expressions for the evaluation of G^0. By introducing the Bloch functions $|n\vec{k}\rangle = |n\vec{k}\rangle$ as a complete set, G^0 can be written in the form

$$G^0(E^+) = \sum_{n\vec{k}} \frac{|n\vec{k}\rangle\langle n\vec{k}|}{E^+ - E_{n\vec{k}}}. \quad (36)$$

Defining the spectral density operator $A^0(E)$ (Ref. 31) by

$$A^0(E) = \sum_{n\vec{k}} |n\vec{k}\rangle\langle n\vec{k}|\delta(E - E_{n\vec{k}}), \quad (37)$$

and using the Dirac identity

$$\lim_{\epsilon\to 0^+}\int \frac{f(t)\,dt}{t+i\epsilon} = \text{P}\int \frac{f(t)\,dt}{t} - i\pi f(0), \quad (38)$$

where P denotes the principal part of the integral, we obtain

$$G^0(E^+) = \text{P}\int \frac{A(E')}{E-E'}\,dE' - i\pi A(E). \quad (39)$$

IV. THE CHOICE OF THE BASIS SET

The selection of the basis set for the expansion of the operators in actual calculations is obviously problem-dependent and must be guided by physical considerations. It also constitutes one of the most important choices that determine the efficiency of the method.

The traditional basis set, i.e., the Wannier functions, have proved very convenient for the formal development of the method.[12,13] For actual calculations, on the other hand, they pose problems because their construction is quite laborious.[14,15] In this work we chose to use a set of LCAO's. Such sets have recently proved to be quite powerful and efficient in band-structure calculations[16,17] and a great deal of experience for their use has been amassed by atomic and molecular calculations.

The first step is to determine an LCAO set that is capable of yielding an accurate charge density and band structure for the perfect crystal (see Sec. V for the actual basis set used). The same set is then used to represent the operators $G^0(E)$ and U for the description of an unrelaxed vacancy. This choice can be justified as adequate for describing the perturbed crystal by invoking the usual quantum-chemical practice of using LCAO's characteristic of particular atoms only at the atomic sites. It is also adequate to describe the perturbation, because U is expected to be essentially the negative of an atomic-like potential. In other words, since the LCAO's at atomic sites in the perfect crystal are adequate for the description of bond formation, the same orbitals at the vacant site and the neighboring sites can be expected to describe the breaking of the bonds.

Our choice of a common LCAO set to describe both the perfect and the perturbed crystal has a number of useful consequences: in particular, by using an orthonormal LCAO set (see Sec. V for details), the operator equations of Secs. II and III can be read directly as matrix equations. The matrix elements between basis orbitals and Bloch functions necessary to evaluate the Green's-function matrix [cf. Eq. (36)] become very simple. Finally, because the perfect-crystal Hamiltonian has a finite spectrum, any ambiguities[9,3] associated with the termination of the energy integral of Eq. (39) are completely removed.

In order to exploit the point-group symmetry of the vacancy, we form symmetrized linear combinations of LCAO's on each shell of atoms surrounding the vacancy (shell orbitals). Finally, since the LCAO's, and therefore the shell orbitals are not orthogonal, we orthogonalize each shell orbital to all shell orbitals closer to the defect site, forming a set of orthonormal shell orbitals (OSO). This procedure, in contrast to the symmetric Löwdin orthogonalization,[30] has the advantage of preserving the localization of the basis functions and allows an easy and systematic study of the contributions of various shells and of convergence with respect to the number of shells. Also, additional shells may be added to the basis without the need of recalculating any of the matrix elements between existing orbitals. The orthonormalization process by itself avoids the need for carrying the overlap matrix in Eqs. (30)–(36), leading to large savings in computer time. The details of the symmetrization and orthogonalization procedure are given in Appendix A.

V. CALCULATIONS

In this section we will describe the calculational procedures used in applying the Green's-function method to study an ideal (undistorted) vacancy in Si. The calculations consist of three major steps: 1) solution of the bulk problem and the calculation of the perfect-crystal Green's function G^0, 2) solution of the Green's-function equations for a given potential, and 3) construction of the self

consistent potential. The methods of calculation employed at each step are described and tested in Secs. V A–C below.

A. Perfect crystal

1. Energy bands

In the self-consistent pseudopotential formalism,[41] the interaction between the valence electrons and the ionic core is modeled by a potential whose short-range form is fitted to the bulk properties,[11] atomic term values,[42] or atomic calculations.[43] The valence charge density, and thereby the valence crystalline potential, is then determined self-consistently within the local-density theory.[44] Usually, such calculations are carried out using a plane-wave basis set. One of the advantages of this set is that convergence studies can be carried out in a straightforward and systematic manner. We have therefore used a plane-wave basis set to calculate fully convergent energy bands for Si to serve as standards against which to compare our LCAO calculations. Such a calculation is necessary because we wish to use a minimal LCAO basis, optimizing the orbital exponents to yield bands that are in good agreement with known accurate calculations. We have carried out plane-wave calculations using the ionic potentials determined in Refs. 42 and 43. Both these potentials are specified in terms of the same analytical expression, namely,

$$V_{\text{ion}}(q) = (a_1 q^2)\{\cos(a_2 q) + a_3\} e^{a_4 q^4}. \tag{40}$$

In Ref. 42, the coefficients in Eq. (40) were fitted to experimental atomic term values, whereas in Ref. 43 they were fitted to atomic calculations. Both sets of coefficients are given in Table I. The charge density for each iteration was calculated using ten special k points of Chadi and Cohen.[45] We found 65 plane waves in the wave-function expansion to be sufficient to obtain a converged expansion of the crystalline charge density. The crystalline potential was calculated using the same exchange correlation as in the atomic calculations used to fit Eq. (40), i.e., the X_α approximation with $\alpha = 0.8$. The band-structure energies for the high-symmetry points are given in the first two columns of Table II. In particular, one should notice that the fundamental band gap is 0.6 and 0.8 eV for the two potentials. The discrepancy with the experimental value of 1.17 eV is attributed to the failure of the local approximation in the Kohn-Sham[44] density-functional formalism. In the following, we will use the ionic potential from Ref. 43, which is derived from atomic calculations. This choice makes our calculations both free from any empirical adjustments, and internally consistent, i.e., every quality has been calculated entirely within the local-density theory with the same exchange. In our defect calculations we use the LCAO's of Chadi[37] which consist of s, p, and d orbitals, and one f orbital transforming as the potential (xyz). We have carried out fully self-consistent calculations for silicon in this basis without any shape approximation for the crystalline potential.[46] The band energies of the LCAO calculation with an LCAO self-consistent potential are compared with the fully converged plane-wave calculations in Table II. The largest discrepancy between the two calculations is about 0.2 eV, in the band-gap region, i.e., the same as obtained by Chadi[37] using a different potential. In particular, the band gap is 1.0 eV.

TABLE I. The values (in atomic units) of parameters defining the ionic Si pseudopotential V_{ion} [Eq. (41)] and bulk pseudopotential V_{bulk} [Eq. (43)]. The normalization volume for V_{ion} is 135.1 (a.u.)^3.

	$V^{\text{ion}}_{\text{ion}}$ (Ref. 42)	$V^{\text{ion}}_{\text{ion}}$ (Ref. 43)	$V^{\text{bulk}}_{\text{bulk}}$
a_1	−1.12507	−1.233	20.0
a_2	0.79065	0.7370	0.633
a_3	0.35201	−0.3969	−17.7
a_4	−0.01807	−0.0177	0.459

2. Green's function

Once the band-structure calculation is carried out, the matrix elements of the Green's operator $G^0(E)$ between any pair of basis orbitals (the OSO's) can be evaluated using Eqs. (37) and (39). For a point defect of T_d symmetry, as is the case of the unrelaxed vacancy, the $\vec{k}$-space summation has to be done only over one irreducible wedge W, with the summand chosen by symmetry considerations so that the result is equivalent to a full-zone summation. If we denote each OSO by $|\alpha i l\rangle$, where α is the representation, i is a running index for OSO's, excluding partners, and l is a partner index, it is straightforward to show that

$$\langle \alpha i l | A(E) | \alpha' i' l' \rangle = \frac{48}{d_\alpha} \operatorname{Re}\left(\sum_n \sum_{\vec{k}} \sum_{i''} \langle \alpha i l | n\vec{k} \rangle \langle n\vec{k} | \alpha' i'' l'' \rangle \times {}^{*}(E - E_n(\vec{k}))\, \delta_{\alpha\alpha'} \delta_{ll'} \right), \tag{41}$$

where d_α is the dimensionality of the αth representation. It is $\delta_{\alpha\alpha'}$ and $\delta_{ll'}$ in Eq. (41), which arise from symmetrization of the basis orbitals, that provide the savings in labor. The evaluation of the matrix elements of $A(E)$ as given by Eq. (41) was carried out using the Gilat-Raubenheimer[47,48]

TABLE II. Band-structure energies for the high-symmetry points using potentials from Ref. 42 and Ref. 43. The point along the Δ axis is $\vec{k}=(0.85,0,0)$. The number of plane waves corresponds to an energy cutoff $\vec{k}+\vec{G}=9$ Ry. The Slater orbitals are those of Chadi (Ref. 37). The band energies for those orbitals have been shifted by 0.14 eV with respect to the plane-wave results (see Ref. 37).

	Plane waves[a]	Plane waves[b]	Slater orbitals[b]	Deviation from plane-wave[b] results
Γ_{1v}	−12.79	−12.89	−12.52	−0.37
$\Gamma_{25'v}$	0.00	0.00	0.00	
$\Gamma_{15'c}$	2.88	2.98	3.10	−0.12
$\Gamma_{2'c}$	3.49	3.72	3.92	−0.2
X_{4v}	−8.48	−8.63	−8.38	−0.25
X_{1v}	−3.14	−3.06	−2.95	−0.11
X_{1c}	0.78	0.95	1.11	−0.16
X_{4c}	11.83	11.67	11.93	−0.26
$L_{2'v}$	−10.39	−10.55	−10.25	−0.3
L_{4v}	− 7.52	− 7.52	− 7.27	−0.25
L_{1v}	− 1.36	− 1.34	− 1.29	−0.21
L_{1c}	1.60	1.38	1.59	−0.21
L_{3c}	3.69	3.89	4.07	−0.13
Δ_{1v}	− 9.63	− 9.76	− 9.50	−0.26
Δ_{2v}	− 7.21	− 7.38	−7.12	−0.26
Δ_{5v}	− 3.04	− 2.97	− 2.86	−0.11
Δ_{1c}	0.62	0.78	0.97	−0.19
Δ_{2c}	1.20	1.39	1.56	−0.17

[a] Potential from Ref. 42.
[b] Potential from Ref. 43.

technique. Convergence tests were carried out and it was found that 308 $\vec{k}$ points in the irreducible wedge were adequate. Once the matrix elements of $\Delta(E)$ were calculated, the matrix elements of $G^0(E)$ were obtained using Eq. (39). The large number of Hilbert transforms of rather strongly varying functions with sharp peaks has been handled using a fast-Hilbert-transform (FHT) algorithm developed by Cooley and Bernholc.[43] The use of the FHT resulted in a very accurate determination of the Green's-function matrix elements on an energy mesh spaced by 0.07 eV. This mesh was found dense enough to allow interpolation for intermediate energy values when needed. The Green's-function matrix elements were calculated once and stored.

B. Solutions for a given defect potential (the non-self-consistent vacancy)

For the purpose of testing our computational scheme of determining the solution for a given defect potential, we first used the negative of a bulk self-consistent "atomic" pseudopotential, obtained by decomposing the full crystalline pseudopotential into a sum of atom-centered spherical pseudopotentials, neglecting nonspherical terms. The resulting atomic pseudopotential was fitted to an analytical expression of the form

$$V_{\text{atom}}(r)=a_1e^{-a_2r^2}+a_3e^{-a_4r^2}, \tag{42}$$

The a_i's are given in Table I. Superposition of these atom-centered spherical potentials reproduces the total bulk pseudopotential quite adequately (Fig. 1). Thus, using the negative of one of these atom-centered potentials as a model defect potential corresponds to an approximation for the vacancy that amounts to removing an atom from the

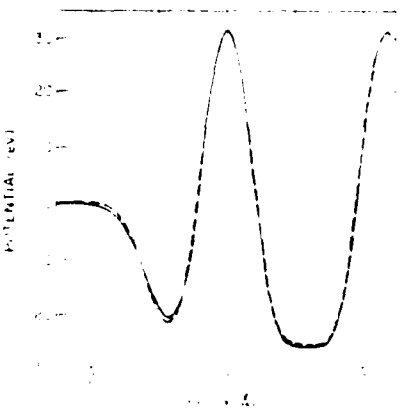


FIG. 1. Comparison of the truly self-consistent bulk silicon potential (solid line) with the superposition of atom-centered spherical potentials (dashed line) along the bonding (or antibonding) direction.

crystal without allowing the remaining valence electrons to redistribute (absence of self-consistency). The calculations using the non-self-consistent vacancy potential, as defined above, serve two purposes. First, they allow us to test the convergence properties of our computational procedure for a given defect potential, and, second, to assess the importance of self-consistency and the resulting screening fields once the fully self-consistent solutions are obtained (Sec. V C below).

Given the defect potential, its matrix elements between pairs of orthogonalized shell orbitals (OSO's) were calculated numerically on a cubic mesh. Symmetry was again used to reduce the integration region by a factor of 24. We found that 1000–2000 points per atom in the irreducible part of the cube were sufficient for an accuracy of 2–3 mRy in the potential matrix elements.[50] Having the matrix elements of both $G^0(E)$ and the defect potential U, the quantity $D(E)$ given by Eq. (9) is evaluated in a straightforward way. Making use of symmetry once more, one gets[18,19]

$$D(E) = \prod_{\alpha} [D_{\alpha}(E)]^{d_{\alpha}} , \tag{43}$$

where $D_{\alpha}(E)$ is defined as in Eq. (9), but evaluated in the subspace of OSO's of the αth representation. Accordingly, bound states in the gap belonging to the αth representation are obtained by locating the zeros of $D_{\alpha}(E)$. Similarly, one can show that

$$\Delta N(E) = \sum_{\alpha} \Delta N_{\alpha}(E) , \tag{44}$$

where $\Delta N_{\alpha}(E)$ is again defined as $\Delta N(E)$, but within the subspace of OSO's of the αth representation (ΔN_{α} is defined so as to contain the degeneracy factor d_{α}).

The calculations were carried out using one, two, and three shells of orbitals (one shell means orbitals at the vacant site only). In all cases, we find only one bound state of T_2 symmetry. Its energy level is at 0.75, 0.75, and 0.76 eV for one, two, and three shells, respectively, indicating an adequate convergence with respect to the number of shells. The corresponding changes in state densities within the band continua, shown as dashed curves in Fig. 2, were also found to be indistinguishable for one, two, and three shells on the scale of the figure. Only changes in densities of states of A_1 and T_2 symmetries are shown. Changes in the densities of states of the other symmetries are of order 0.1 and thus insignificant by comparison. These results are interpreted in Sec. VI, where comparison is made with fully self-consistent results.[11] Integrals of the state-density changes provide an additional check of the calculations. It can easily be shown that Levinson's theorem [Eqs. (22) and (23)] is valid for states of each symmetry separately. Thus, the state-density changes for each symmetry ought to integrate to the number (integer) of bound states of the same symmetry. Indeed, we found that the change in the density of T_2 states integrates to -6, compensating the sixfold degenerate T_2 bound state in the fundamental gap. The changes in the densities of states of the remaining symmetries integrate to zero, as they should. Thus, since the creation of a neutral vacancy involves the removal of four electrons, we conclude that the T_2 bound state contains two electrons.

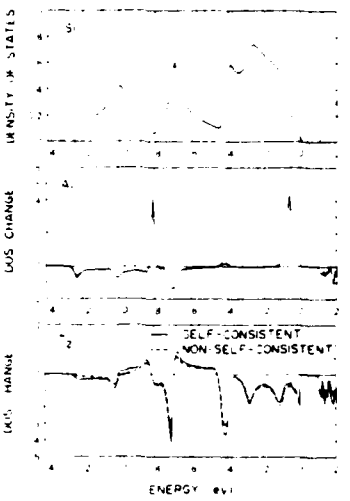


FIG. 2. The density of states of silicon (top panel) and the change in the density of states of A_1 and T_2 symmetries. The curves are broadened by 0.2 eV and the reference energy is the top of the valence bands.

C. Fully self-consistent solutions

After the calculations have been initiated with a starting potential (e.g., the non-self-consistent potential described above or some other convenient form), the Green's-function equations have to be iterated to self-consistency. At every iteration the construction of a new potential consists of the following steps: (1) calculation of the change in charge density in the valence bands and of the charge density associated with the bound states, (2) solution of the Poisson equation, and (3) calculation of the exchange-correlation term within the local-density theory[11].

The change in the charge density associated with the valence bands was obtained by carrying out the matrix integral (35). Since the integrand is a rather singular function (see below), the integral was calculated by the adaptive variable-step size

trapezoidal rule. The average step length was 0.07 eV and was monitored by examining $\mathrm{Tr}\Delta\rho(E)$, which is equal to the change in the density of states at that energy. The change in the charge density in real space is now

$$\Delta\rho^{\mathrm{val}}(\mathbf{r}) = \sum_{\alpha\alpha'} \psi^{*}_{\alpha}(\mathbf{r})\Delta\rho^{\mathrm{val}}_{\alpha\alpha'}\psi_{\alpha'}(\mathbf{r}) , \tag{45}$$

The charge density associated with the two electrons in the bound state was obtained from the bound-state wave function within the region spanned by Eq. (30), and the normalization constant was calculated from relation (34). It will be shown in the next section by explicit construction that the rather extended tail of the bound-state wave function is canceled by the charge redistribution within the valence bands localizing the total change in the charge density almost entirely within the cavity defined by the nearest neighbors. This localization is important to the efficiency of the method and has been exploited in the next stages of the calculation.

The Hartree part of the defect potential V_H can now be obtained by solving the Poisson equation

$$\nabla^2 V_H = -8\pi\Delta\rho , \tag{46}$$

with the boundary condition, for large r,

$$V(r) \sim -Z/r , \tag{47}$$

where

$$Z = -\int d^3r\,\Delta\rho(\mathbf{r}) . \tag{48}$$

In order to facilitate a numerical solution, it is convenient to add and later subtract a neutralizing, positive, and spherically symmetric charge density ρ^{+}. The charge

$$\rho^{N} = \Delta\rho + \rho^{+} \tag{49}$$

is then neutral and the corresponding potential V^{N} behaves as const $\times r^{-(l+1)}$ at large distances where l is the lowest multipole component of the defect potential allowed by symmetry ($l = 3$ for the T_d group). V^{N} is determined by solving Poisson's equation in a large cube (cube side = 18 a.u.) around the vacancy. Because of the smallness of the r^{-4} and higher-order terms on the surface of the cube, the boundary condition is taken to be $V^{N} = 0$. This geometry allows us to use the very efficient fast-Fourier-transform[51] techniques in order to solve the Poisson equation numerically,[48,49,52] the only constraint being that the numerical mesh must be equally spaced (see Appendix B for the discussion of the algorithm). This algorithm defines the cubic mesh for the calculation of the potential matrix elements (cf. Sec. V B and Ref. 50). The Hartree potential is obtained by subtracting the potential V^{+} corresponding to the charge density ρ^{+}. In our calculations, we choose

$$\rho^{+}(\mathbf{r}) = \frac{1}{\pi^{3/2}}\frac{Z}{r_1^3}\exp(-r^2/r_1^2) , \tag{50}$$

whereby

$$V^{+}(\mathbf{r}) = \frac{2Z}{r}\,\mathrm{erf}\left(\frac{r}{r_1}\right) . \tag{51}$$

Finally, the exchange-correlation potential is calculated in the local-density approximation. Because of the $\rho^{1/3}$ dependence of this potential on the charge density, it is necessary to calculate both the perfect-crystal charge density $\rho^{0}(r)$ as well as $\Delta\rho(r)$ at each mesh point. The defect exchange-correlation potential is then proportional to $[\rho^{0}(r)+\Delta\rho(r)]^{1/3} - [\Delta\rho(r)]^{1/3}$. The calculation is done numerically[50] on the same mesh points used for the Hartree part of the potential.

VI. RESULTS AND DISCUSSION

In this section we present the results of our self-consistent calculations for the unrelaxed vacancy in Si. The calculations were brought to self-consistency using one, two, and three shells for the purpose of checking convergence of charge densities and wave functions with respect to the number of shells.

A. Energy levels and state densities

As with the non-self-consistent calculations described in the previous section, we obtain one bound state in the gap of T_2 symmetry. Its energy level was found to be at 0.63, 0.66, and 0.68 eV for one, two, and three shells, respectively. Convergence is therefore somewhat slower, but still quite adequate. We note that self-consistency has lowered the position of the bound state by approximately 0.1 eV. This lowering is the result of screening arising from the response of the valence electrons to the removal of the atom, which reduces the strength of the defect potential. The final self-consistent vacancy potential is shown in Fig. 3 and compared with the non-self-consistent vacancy potential defined in Sec. V. Note that the effect of self-consistency is very small, as reflected by the lowering of the bound-state energy level by less than 0.1 eV.

The changes in the densities of states of A_1 and T_2 symmetries are shown as solid curves in Fig. 2. On the scale of the figure, the curves obtained from the one-, two-, and three-shell calculations are indistinguishable. Changes in the densities of states of the other symmetries are again insignificant by comparison. The change in the densities

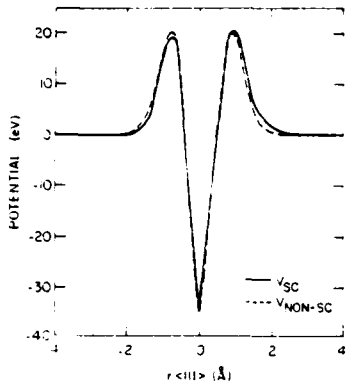


FIG. 3. Comparison of the fully self-consistent vacancy potential to the negative of the spherical atom-centered potential extracted from bulk data. See text.

of T_2 states integrates to -6, as required by Levinson's theorem, and the changes in the densities of states of the remaining symmetries integrate to zero. The self-consistent state-density changes are compared in Fig. 2 with the corresponding changes obtained without self-consistency, indicating once more that the effect of self-consistency is small.

In Fig. 2 we have included a plot of the density of states of the perfect crystal (top panel), against which the calculated changes induced by the vacancy can be compared. We note that a sharp resonance of A_1 symmetry appears at -8 eV, where the perfect-crystal state density has a minimum. The antiresonances appear at or near maxima in the perfect-crystal state densities. These results are consistent with the general analysis given in Sec. II B, since extrema in the perfect-crystal state density correspond to critical points in the band structure. The A_1 resonance at -0.7 eV, on the other hand, is caused by the particular nature of the vacancy potential and will be discussed further later on.

B. Wave functions and charge densities

For further analysis of the nature of the solutions we examine wave functions and charge densities. The most convenient and informative way to display these results is in the form of contour maps in a (110) plane. All the results that will be presented were obtained from three-shell calculations. The three-shell basis is sufficiently large to allow plotting beyond the second neighbors of the vacancy.

In Fig. 4 we show the charge density of the perfect crystal, the charge density in a crystal containing a single vacancy, and the change in the charge density produced by the introduction of the

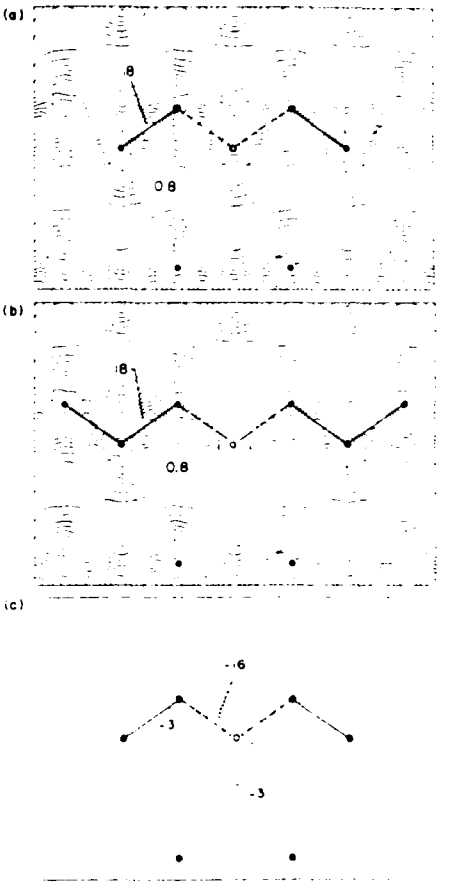


FIG. 4. Contours of constant electron density (in electrons per bulk Si unit cell) for (a) charge density of perfect silicon. (b) Charge density in the presence of the vacancy. (c) The change in the charge density i.e., the difference between (b) and (a).

vacancy. As an internal check of the calculations, the change in the charge density is found to integrate to -4, corresponding to the net number of electrons that have been removed from the crystal. We note that the change in the charge density is localized almost entirely within the cavity defined by the nearest neighbors. This charge produces a potential that becomes equal to $-4e^2/r$ beyond the nearest neighbors and cancels exactly the $4e^2/r$ tail of the ionic part of the vacancy potential. The net vacancy potential is therefore almost completely localized within a distance of a bond length, as we already saw in Fig. 3. These results are consistent with the fact that orbitals on the second and the third shells of atoms do not contribute significantly.

We turn now to examine the wave functions of individual states. In Fig. 5 we show a contour plot of the square of the wave function of the T_2 bound

TABLE III. The integrated changes in the valence-charge density for each representation.

Representation	Charge(electrons)
A_1	−0.10
A_2	0.00
E	−0.04
T_1	−0.03
T_2	−5.20

state. *In contrast to the total change in the charge density* [Fig. 4(c)], *we see that the charge density associated with the bound state is quite extended.* In fact, the basis orbitals on the three shells of atoms included in the plot of Fig. 5 contain only approximately 70% of the total bound-state charge (1.4 bound-state electrons lie inside the three-shell volume). Note that this result does not imply that additional shells of orbitals ought to have been included in the calculation. Recall that an important virtue of the Green's-function method is that it does not require a basis set capable of expanding individual wave functions. The role of the basis set is to represent the defect potential and *change* in the charge density. As we saw already, both of these quantities are quite localized.[33]

The rather delocalized bound-state charge density might at first glance appear to be inconsistent with the fact that the total change in the charge density is highly localized. Note that previous non-self-consistent calculations *assumed* a localized vacancy potential and obtained a delocalized bound state. The present self-consistent calculations confirm this assumption and provide a justification: In addition to the T_2 bound state in the gap, the defect potential induces a series of resonances and antiresonances in the valence bands, which we already saw in Fig. 2. Contour plots of the charge density associated with some of the resonances are shown in Figs. 6 and 7, revealing that the corresponding charge density is quite delocalized. It is a subtle cancellation of the tails of individual states that produces the strong localization of the *net* charge disturbance. Another illustration of this fact is provided by plots of the total change in the charge density for each type of symmetry separately (Fig. 8). Each of them is quite delocalized [the solid contours are the sum of all positive changes (resonances), whereas the dashed contours are the sum of all negative changes (antiresonances)] and not necessarily contained in the volume defined by three shells of atoms. In fact their integrals within the three-shell basis are given in Table III. All five together add up to −5.4. Recall that the bound state contributes only 1.4 electrons in the volume defined by three shells of atoms. The net change within this volume is therefore −4, indicating that indeed the *net* change outside the box is zero and providing an additional internal check of the calculations.

The above results for the total charge densities, the defect-induced change in the total charge density and individual wave functions provide a clear illustration of the advantages of the Green's-function method over cluster methods. Cluster methods attempt to describe the perturbed crystal directly [Fig. 4(b)] and terminate it as a matter of necessity to make the calculation feasible. In contrast, the Green's-function method describes first the infinite perfect crystal [Fig. 4(a)] and then focuses on the change [Fig. 4(c)] which is naturally localized. Furthermore, whereas the Green's-function method needs a basis set capable of de-

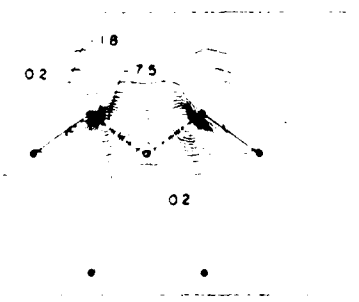


FIG. 5. Contours of constant electron density for the T_2 bound state. Units as in Fig. 4.

FIG. 6. Contours of constant electron density for the A_1 resonances at (a) −5 and (b) −0.7 eV. Units as in Fig. 4.

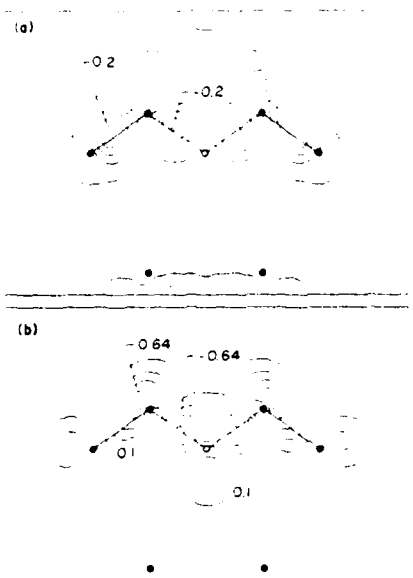


FIG. 7. Contours of constant electron density for the T_2 quasiresonances at (a) −8.4 and (b) −6.8 eV. Units as in Fig. 4.

scribing accurately only the total change in the charge density, which is highly localized [Fig. 4 (c)], the cluster method needs a basis set capable of describing accurately individual wave functions (Fig. 5) which are considerably more extended. Only when the size of the cluster is large enough to contain individual wave functions would a cluster calculation give reliable quantitative results.

C. LCAO analysis

The use of an LCAO basis for the calculations provides a convenient framework for exploring the structure and origins of the solutions. First, we observe that the valence electrons in the perfect Si crystal have mainly s (A_1) and p (T_2) character about each atom. This is a well-known result that has often been exploited in constructing simple

TABLE IV. The decomposition of the bound-state wave function of the undistorted vacancy into the bands of perfect silicon $c_n = \sum_{\vec{k}} |\langle n\vec{k} | \cdots |^2$.

Band no.	c_n (%)
1	0.4
2	4.8
3	23.7
4	51.1
5	10.9
6	4.1
7	2.2
8	1.4
9	0.6
10	0.3

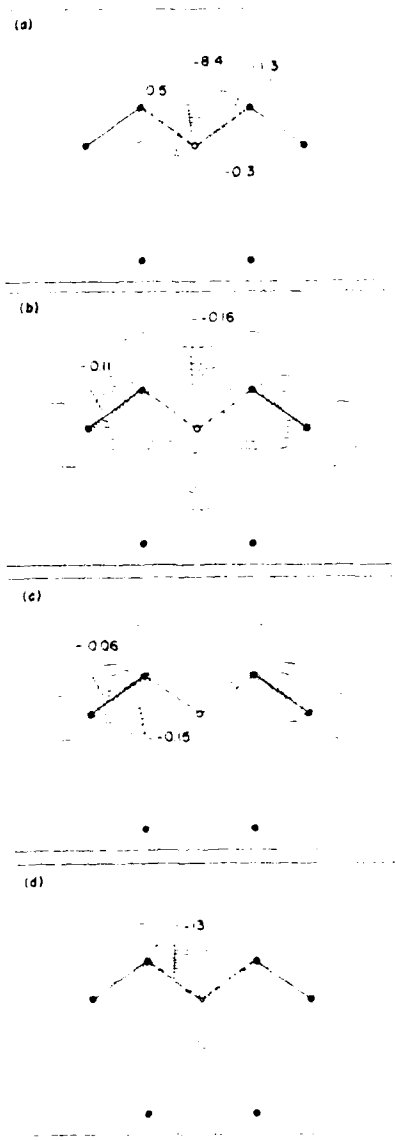
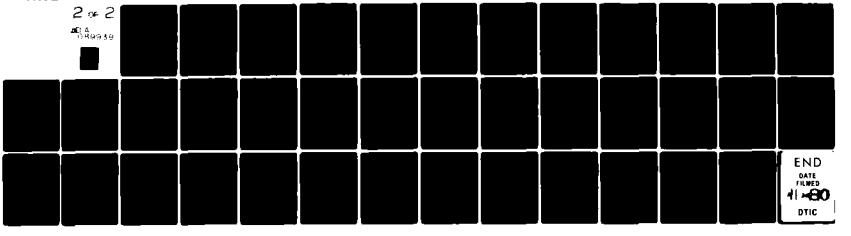


FIG. 8. Contours of constant electron density for the change in the charge density of symmetry (a) A_1, (b) E, (c) T_1, (d) T_2. Units as in Fig. 4.

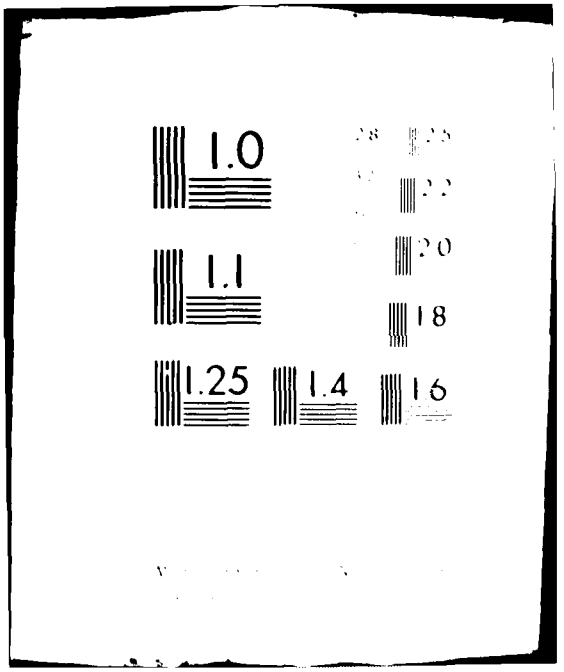
semiempirical tight-binding Hamiltonians.[4] Our self-consistent calculations reveal that the valence electrons also have a small ($\sim E + T_2$) character.[5] Since the creation of a vacancy amounts to removing four valence electrons from the crystal, one can expect that the major changes in the charge and state densities occur in the A_1 and T_2 representations. Our results, Figs. 2 and 3, confirmed this simple LCAO prediction. In particular, note in Fig. 8 that the A_1 and T_2 changes in the charge density are large and concentrated in the cavity defined by the nearest neighbors. The T_1 charge

AD-A089 938 IBM THOMAS J WATSON RESEARCH CENTER YORKTOWN HEIGHTS NY F/G 20/12
THEORETICAL STUDIES OF DEFECTS IN TETRAHEDRAL SEMICONDUCTORS.(U)
AUG 80 S T PANTELIDES, J A VANVECHTEN F49620-79-C-0077
UNCLASSIFIED AFOSR-TR-80-0758 NL

2 OF 2
AD A089938



END
DATE FILMED
11-80
DTIC


1.0
1.1
1.25
1.4
18
16

does not have a p-like shape because it is an average of the "x", "y", and "z" components.) Note also the small change in the E representation (d-like) in Fig. 8(c) in agreement with the LCAO analysis given above. Figure 8(d) shows the change of charge density in the T_1 states. This change, which is also quite small compared with the changes in the charge density of A_1 and T_2 states, comes from linear combinations of orbitals on the nearest neighbors and is therefore localized about those sites. Finally, the change of charge density of A_2 states is too small to be seen on the scale of Fig. 8, because it arises from linear combinations of orbitals on the second-nearest neighbors.

We now focus our attention on the A_1 and T_2 results and show that they can be understood in terms of some very elementary LCAO models. For the perfect crystal, one can use s and p orbitals on every atom and combine them into tetrahedrally directed sp^3 hybrid orbitals. These hybrids can then be combined in pairs to form bonding and antibonding orbitals, which in turn give rise to the valence and conduction bands, respectively.[56] For the crystal containing a single vacancy, one can again use s and p orbitals on every atom and form sp^3 hybrids. Bonding and antibonding orbitals can again be formed by pairing hybrids, with the exception of the four hybrids on the nearest neighbors pointing toward the vacant site. We will refer to these orbitals as "dangling hybrids" (the charge associated with such an orbital is often referred to as a "dangling bond"). As a first approximation, one would expect that the states associated with the vacancy are linear combinations of these four hybrids. This assumption forms the basis of the "defect-molecule" model, originally proposed by Coulson and Kearsley[6] for the vacancy in diamond. Symmetry requires that the four hybrids be combined into an A_1 singlet and a T_2 triplet. In fact, our self-consistent pseudopotential calculations reveal that the T_2 bound state (Fig. 5) and the A_1 resonance at -0.7 eV [Fig. 6(b)] are essentially linear combinations of the dangling-hybrid-like orbitals. The remaining resonances and antiresonances are mainly associated with critical points in the band structure (cf. Secs. II and VI A). The A_1 resonance at about -8 eV appears precisely at the point where the density of states has a cusp. (This cusp arises from the X_1 point in the valence bands and is a peculiarity of the diamond lattice. It may be viewed as a pseudogap, since it is precisely at the point where a gap opens up in the zinc-blende structure.) In fact, this resonance is a band-structure effect and appears at the same place for several substitutional acceptors.[57] The contour plot of the charge redistribution corresponding to this resonance is shown in Fig. 6(a). Note that it is quite localized since it is essentially a bound state. It consists mainly of s-like orbitals on the nearest neighbors.

In Fig. 7 we show the changes in the charge density corresponding to the two T_2 quasiresonances at (a) -8.4 eV and (b) -6.8 eV, as obtained from the self-consistent calculations. The quasiresonance at -8.4 corresponds to a strengthening of the back bonds [recall however, that, overall, the back bonds are not strengthened because the *total* change is localized entirely within the cavity defined by the nearest neighbors; cf Fig. 4(c).] The quasiresonance at -6.8 eV shifts some of the charge into the antibonding direction in the normally empty interstitial regions.

D. The effective-mass nature of the bound state

In the self-consistent Green's-function calculations, we obtain the bound-state wave functions in terms of our LCAO basis set in the form

$$\psi(\vec{r}) = \sum_{\mu} C_{\mu}\phi_{\mu}(\vec{r}), \tag{52}$$

where ϕ_{μ} is an OSO (for simplicity, we use $\mu = \{\alpha, i, l\}$ as a composite index for an OSO). A common expansion of the ψ, e.g., in the case of effective-mass theory, is in terms of Bloch functions in the form

$$\psi(\vec{r}) = \sum_{n\vec{k}} F_n(\vec{k})\psi_{n\vec{k}}(\vec{r}). \tag{53}$$

Combining (52) and (53), we immediately obtain

$$F_n(\vec{k}) = \langle n\vec{k}|\psi\rangle = \sum_{\mu} C_{\mu}\langle n\vec{k}|\phi_{\mu}\rangle. \tag{54}$$

An alternative expression can be obtained by using Eqs. (7) and (36):

$$F_n(\vec{k}) = \sum_{\mu} C_{\mu}\langle nk|U|\phi_{\mu}\rangle/(E - E_{n\vec{k}}). \tag{55}$$

This expression shows that only those OSO's for which the matrix elements of U are nonzero need be kept in calculating $F_n(k)$. We have calculated these quantities and summed them over the Brillouin zone for each band in order to establish the relative weight of each band in the expansion of ψ. The results are given in Table IV. We note that 80% of the wave function comes from the top three valence bands, with the remainder distributed over a total of seven more bands. The lowest conduction band contributes 11%, whereas the rest of the conduction bands together contribute less than 9%.

In Fig. 9 we give plots of the various $F_n(\vec{k})$ functions along two important symmetry directions in $\vec{k}$ space. Only the $F_n(\vec{k})$ for bands 2, 3, 4, and 5 can be plotted on the scale of Fig. 9. This figure

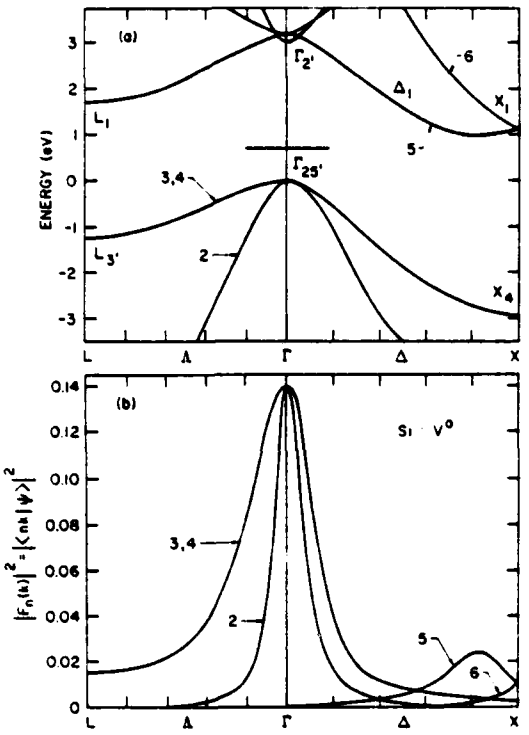


FIG. 9. The band structure of silicon of the vicinity of the band gap and the envelope functions $F_n(\vec{k})$ plotted along the Λ and Δ directions. See text.

illustrates clearly that ψ is composed mainly from Bloch functions in the vicinity of the valence-band maximum at $\vec{k}=0$, very much like effective-mass-like bound states. We may recall at this stage that effective-mass equations have recently been rederived[58] for an arbitrary impurity potential by making only two approximations: (1) restricting the number of bands that participate in the expansion of the bound-state wavefunction, and (2) evaluating all matrix elements to order k^2 (thus including the so-called umklapp terms[59]). The results shown in Fig. 9 for the vacancy bound-state wave function indicate that the new generalized effective-mass equations might be applicable even for quite deep levels. Numerical work would have to be carried out to establish their actual usefulness. It should be noted, in any case, that the effective-mass equations would be useful only if the defect or impurity potential were known (the theory cannot determine the potential self-consistently) and then only for the bound state(s) in the gap. Further discussion of the connections with EMT is given in Ref. 60.

E. Comparison with tight-binding models

So far, we have carried out Green's-function calculations of three different degrees of complexity: (1) using a semiempirical tight-binding Hamiltonian, reported in paper I,[28] (2) using a non-self-consistent vacancy pseudopotential taken to be the negative of a bulk atomic pseudopotential, and (3) using a fully self-consistent vacancy pseudopotential. (In the last two cases, the underlying band structure is also based on self-consistent pseudopotentials.) We have already seen in Sec. V B that the results of calculations (2) and (3) differ only in their details. We therefore turn now to a comparison with the tight-binding results of Ref. 28. In Fig. 10 we compare the tight-binding results for the changes in the densities of A_1 and T_2 states with the results of the fully self-consistent calculations described above. The two curves in each case are remarkably similar, confirming the usefulness of the tight-binding calculations. We may recall that in the tight-binding model[28] all Hamiltonian matrix elements are assumed to be identical to those of the unperturbed bulk, even those corresponding to the backbonds. The self-consistent results for the total change in the charge density [Fig. 4(c)] show that changes are restricted within the vacancy cavity and provide an explanation and justification for the validity of the tight-binding model.

The bound-state energy, however, obtained by the tight-binding calculations[28] is far too low (0.3 eV compared with the present 0.7 eV). We speculate that this unsatisfactory result is a consequence of the fact that the tight-binding conduction bands are not very accurate. As discussed in Ref. 28, the valence-band state-density changes are deter-

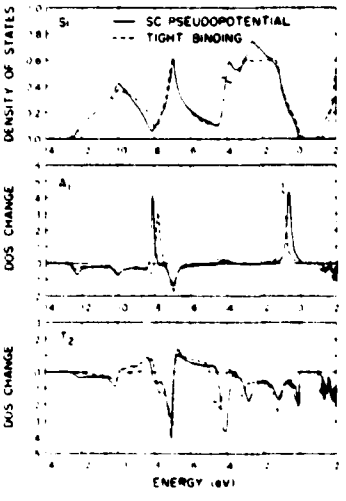


FIG. 10. Comparison of the tight-binding and self-consistent results for density of states (DOS) of perfect silicon and for the DOS changes of A_1 and T_2 symmetries. The curves are broadened by 0.2 eV and the reference energy is the top of the valence bands.

mined primarily by the valence bands, whereas the position of the level in the gap is determined by a subtle balance between valence- and conduction-band contributions. Note that this observation [Eq. (43) of Ref. 28] is not in contradiction with the fact that the bound-state wave function is composed mainly of valence-band Bloch functions (Fig. 9). The tight-binding result comes about because the bound-state wave function is expanded in terms of s and p LCAO's which give rise to both valence and conduction bands. Note that Eq. (43) of Ref. 28 is only a condition that determines the position of the bound-state energy level and contains no information about the composition of the bound-state wave function. On the other hand, this relation suggests that the bound-state energy level obtained by the tight-binding model is too low because the tight-binding conduction bands are too narrow compared with the true sp conduction bands, so that the integral on the right-hand side of Eq. (43) of Ref. 28 is larger than it should be.

VII. SUMMARY AND CONCLUSIONS

In paper I, we have shown that the Koster-Slater Green's-function technique becomes very efficient and accurate when the Wannier functions of the original formulation are replaced by an LCAO basis set. In that paper the method was applied to tight-binding Hamiltonians and its advantages over small- and large-cluster techniques were explicitly illustrated.

In this paper we have demonstrated the feasibility of the Green's-function method in conjunction with first-principles potentials and band structures and we have extended it to fully include self-consistency effects. The principal advantages of this method are the exploitation of (1) the short-range nature of the defect potential, (2) the translational symmetry of the host crystal, and (3) the analytic separation between the bulk and defect properties. Our first application of the method has been to the vacancy in Si. We have found one bound threefold degenerate state of T_2 symmetry at 0.7 eV, which is occupied by only two electrons. There are also several resonances and antiresonances within the valence bands which are associated with the vacancy. The wave functions associated with the bound state, the resonances, and the antiresonances are individually quite *extended*, while the vacancy potential is *short ranged* and extends only up to the nearest neighbors of the vacancy. The localization of the potential is caused by a detailed cancellation of the bound-state charge density by a corresponding charge redistribution within the valence bands.

ACKNOWLEDGMENTS

It is a pleasure to thank S. G. Louie and A. R. Williams for valuable discussions. We would also like to thank J. W. Cooley for advice concerning the use of fast-Fourier-transform algorithms. This work was supported in part by AFOSR under Contract No. F49620-77-C-0005.

APPENDIX A: SYMMETRIZATION AND ORTHOGONALIZATION PROCEDURE

The symmetrized combinations of LCAO's on each shell surrounding the vacancy can be generated by letting the projection operator

$$P^{(j)}_{\mu\mu} = \frac{l_j}{h} \sum_R \Gamma^{(j)}(R)^*_{\mu\mu} P_R \tag{A1}$$

for each representation j act on each of the basis functions on that shell. [R is a rotation belonging to the point group, P_R is the corresponding rotation operator, l_j is the dimension of the irreducible representation, and $\Gamma^{(j)}(R)$ is the matrix of the representation for that rotation.]

The above procedure has been automated and both the symmetrized shell orbitals and the OSO's have been generated on a computer. The result of applying $P^{(j)}_{\mu\mu}$ on an LCAO function is

$$P^{(j)}_{\mu\mu}\phi_\alpha(\vec{r}-\vec{\tau}) = \frac{l_j}{h} \sum_R \Gamma^{(j)}(R)_{\mu\mu}\phi_\beta(\vec{r}-\vec{\tau}'), \tag{A2}$$

where α denotes the transformation property of an LCAO (s, x, etc.), $\vec{\tau}$ is its position; $\beta = P_R^{-1}\alpha$ and $\tau' = P_R\tau$ denote the transformation property and the position of the rotated orbital, respectively. After discarding the linearly dependent functions, the partners to the symmetrized shell orbitals are formed with the help of the transfer operator $P^{(j)}_{\lambda\mu}$. Finally, the resulting basis functions are orthonormalized by the Schmidt process. The orthonormalized shell orbitals (OSO) can then be compactly expressed in terms of the LCAO's on the shell in question and the OSO's on the previous shells.

APPENDIX B

In this appendix we will solve the Poisson equation

$$\nabla^2 V = -8\pi\rho, \tag{B1}$$

subject to the boundary condition zero applied on the surface of a cube. Let us denote the cube side by a. Since V vanishes on the surface of the cube, V can be expanded in a three-dimensional sine series. If the charge density ρ is given on a cubic mesh of N points in each direction, the expansions

$$V(x_1,x_2,x_3)=\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\sum_{k_3=1}^{N} d(k_1,k_2,k_3)\sin\frac{\pi k_1 x_1}{Na}\times\sin\frac{\pi k_2 x_2}{Na}\sin\frac{\pi k_3 x_3}{Na} \quad \text{(B2)}$$

and

$$\rho(x_1,x_2,x_3)=\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}\sum_{k_3=1}^{N} c(k_1,k_2,k_3)\sin\frac{\pi k_1 x_1}{Na}\times\sin\frac{\pi k_2 x_2}{Na}\sin\frac{\pi k_3 x_3}{Na} \quad \text{(B3)}$$

are related by

$$d(k_1,k_2,k_3)=-\frac{8N^3a^2}{\pi(k_1^2+k_2^2+k_3^2)}\,c(k_1,k_2,k_3)\,. \quad \text{(B4)}$$

The Fourier analysis to determine $c(k_1,k_2,k_3)$ and the Fourier synthesis to obtain $V(x_1,x_2,x_3)$ are carried out iteratively using the one-dimensional sine-transform algorithm.

APPENDIX C

In this appendix we discuss the similarities and differences between our implementation of the Green's-function formalism and that developed independently by Baraff and Schlüter (BS).[30] The two formulations have thus far produced virtually identical results for the unrelaxed vacancy in Si.

Our approach is based directly on operator equations which are general results of scattering theory. The basic operators are $G^0(E)$ and U. All quantities of physical interest are given as traces or determinants of operators that involve only $G^0(E)$ and U. (The trace and the determinant of an operator are invariant in the sense that they can be calculated by using any orthonormal complete set of functions to represent the operator in matrix form.) In our work, we chose a basis set which is physically transparent and, at the same time, practical for accurate self-consistent calculations.

BS's formulation, on the other hand, is based on expansions of wave functions and the Green's function in two different basis sets. Expressions for the quantities of interest are *derived* in terms of the expansion coefficients. The same expressions can, however, be obtained directly from the standard scattering-theoretic operator equations by representing the operators in the two basis sets. Differences arise in the final expressions of the two approaches largely because of the two different basis sets as opposed to our single set. In what follows we compare our choices with those of BS and identify advantages and disadvantages.

The central operator of our formulation is

$$Q_1(E)=1-G^0(E)U\,, \quad \text{(C1)}$$

For example, bound-state energies are given by the zeros of the determinant of $Q_1(E)$. In contrast, the corresponding BS result is equivalent to using the operator

$$Q_2(E)=U-UG^0(E)U\,. \quad \text{(C2)}$$

We observe immediately that

$$Q_2(E)=UQ_1(E)\,, \quad \text{(C3)}$$

so that, as long as $\det\|U\|\neq 0$, either Q_1 or Q_2 will in principle give identical results. In practice, however, differences can in fact arise.

One of the appealing properties of $Q_2(E)$ is that it satisfies a variational principle, in the sense that it yields energies that are accurate to second order in the wave function.[30] This property suggests that a desired degree of accuracy can be achieved with fewer basis orbitals. However, it is straightforward to see that, as long as one uses the same number of orbitals to represent U and G^0 in matrix form, the factorization (C3) follows. As a consequence, Q_1 and Q_2 will have precisely the same zeros even though one is variational and one is not. In order to take advantage of the variational nature of Q_2 one must use two different-size basis sets for U and G^0. The factorization (C3) is then not possible and Q_1 and Q_2 would yield different results. In particular, BS used a set of LCAO's at each atomic site, including the vacant site, for G^0 (the "inner" set), and a smaller set that did not include orbitals at the vacant site for U (the "outer" set). The variational principle was then in effect. In contrast, we included orbitals at the vacant site for both G^0 and U, so that the use of Q_1 and Q_2 would yield identical results. We believe that including orbitals at the vacant site for both U and G^0 is mandated by physical considerations, even though the size of the U matrix is increased. As we saw in Sec. IV, the role of the basis set is to represent U and the change in the charge density $\Delta\rho$, both of which are highly localized in the vacant atomic volume (Figs. 3 and 4). In fact, one can argue[53] that orbitals at the vacant site alone ought to be sufficient to expand $\Delta\rho$. We further believe that the omission of orbitals at the vacant site by BS has affected their results, albeit in a minor way. The effect is noticeable for states of A_1 symmetry which have nonzero amplitude at the origin. In particular, we find an A_1 resonance at $E_v-0.7$ eV, whereas BS find it at $E_v-1.1$ eV. We believe that the most likely source of this discrepancy is BS's omission of orbitals at the vacant site in their outer set, i.e., the set that is used to expand the wave functions and hence $\Delta\rho$ and U.

The use of Q_2 with two different-size basis sets has an additional disadvantage[51] that must be dealt with carefully: Note that $\det\|Q_2\|$ is in principle

proportional to $\det\|U\|$. Note also that the strength of the method lies in the fact that the matrix of U is nonzero in a small subset [cf. Eq. (26)]. Normally, one would like to increase this subset and check the convergence of the results. As convergence is reached, $\det\|U\|$ tends to zero and hence $\det\|Q_2\|$ tends to zero as well. Since Q_2 is inverted in the calculation of the change in the charge density, one must exercise caution in carrying out convergence tests (in the limit of full convergence, when $\det\|U\|=0$, special procedures must be used to ensure the inversion of Q_2 is effected in the subspace in which $\det\|U\|$ is nonzero). None of these complications arise when one works with Q_1.

Finally, one last technical difference between our approach and that of BS is the following: As stated earlier, we work with operator equations and represent each operator in matrix form so that the operator equations become matrix equations.[62] Thus, we in fact calculate the Green's-function matrix in the chosen basis set. BS, on the other hand, as they point out, do not calculate matrix elements of the Green's function in any particular set of states. Instead, they *expand* the Green's function in a set of states as follows:

$$G^0(E,r,r')=\sum_{mm'}\Phi_m^*(r)\bar{G}^0_{mm'}(E)\Phi_{m'}(r')\,. \tag{C4}$$

The $\bar{G}^0_{mm'}(E)$ are expansion coefficients which are evaluated by expanding the Bloch functions in terms of the orbitals $\Phi_m(r)$. [Note that the Bloch functions are calculated in a plane-wave basis and then *fitted* to an expansion in terms of the $\Phi_m(r)$.] In fact, one can show that the $\bar{G}^0_{mm'}(E)$ would be Green's-function matrix elements if the Φ_m were orthogonal. If not, and S_{ml} is the overlap matrix of the Φ_m, we have

$$\bar{G}^0_{mm'}(E)=\sum_{ll'}S^{-1}_{ml}G^0_{ll'}(E)S^{-1}_{l'm'}\,, \tag{C5}$$

where the $G^0_{ll'}(E)$ are true Green's-function matrix elements and can be evaluated in a straightforward way as described in Sec. V.

[1] W. Kohn, *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1957), Vol. 5.
[2] S. T. Pantelides and C. T. Sah, Phys. Rev. B 10, 621 (1974); 10, 638 (1974); J. Bernholc and S. T. Pantelides, *ibid.* 15, 4935 (1977).
[3] A. Baldereschi and N. Lipari, in *Proceedings of the Thirteenth International Conference on the Physics of Semiconductors*, edited by F. G. Fumi (Marves, Rome, 1977), p. 595.
[4] M. Altarelli, W. Y. Hsu, and R. A. Sabatini, J. Phys. C 10, L605 (1977); S. T. Pantelides, Solid State Commun. 30, 65 (1979).
[5] For a recent review, see S. T. Pantelides, Rev. Mod. Phys. 50, 797 (1978).
[6] C. A. Coulson and M. J. Kearsley, Proc. R. Soc. London A 241, 433 (1957).
[7] G. D. Watkins and R. P. Messmer, Phys. Rev. Lett. 25, 656 (1970); Phys. Rev. B 7, 2568 (1973).
[8] C. Weigel, D. Peak, J. C. Corbett, G. D. Watkins, and R. P. Messmer, Phys. Rev. B 8, 2906 (1973).
[9] F. P. Larkins, J. Phys. C 4, 3065 (1971); 4, 3077 (1971).
[10] B. Cartling, B. Roos, and U. Wahlgren, Chem Phys. Lett. 21, 380 (1973); B. Cartling, J. Phys. C 8, 3171 (1975); 8, 3183 (1975).
[11] G. D. Watkins and R. P. Messmer, in *Computational Methods for Large Molecules and Localized States in Solids*, edited by F. Herman, A. D. McLean, and R. K. Nesbet (Plenum, New York, 1972), p. 133.
[12] S. G. Louie, M. Schlüter, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. B 13, 1654, (1976).
[13] J. D. Joannopoulos and E. J. Mele, Solid State Commun. 20, 729 (1976).
[14] E. Kauffer, P. Pêcheur, and M. Gerl, J. Phys. C 9, 2913 (1976); Phys. Rev. B 15, 4107 (1977).
[15] See M. Lifshitz, Nuovo Cimento Suppl. 3, 716 (1956) for the original Russian references.
[16] M. Lax, Phys. Rev. 94, 1391 (1954).
[17] G. F. Koster and J. C. Slater, Phys. Rev. 94, 1392 (1954); 95, 1165 (1954).
[18] J. Callaway, J. Math. Phys. 5, 783 (1964); Phys. Rev. 154, 515 (1967).
[19] J. Callaway and A. Hughes, Phys. Rev. 156, 860 (1967); 164, 1043 (1967).
[20] N. J. Parada, Phys. Rev. B 3, 2042 (1971).
[21] J. Callaway, Phys. Rev. B 3, 2556 (1971).
[22] S. P. Singhal, Phys. Rev. B 4, 2497 (1971); 5, 4203 (1972).
[23] M. Lannoo and P. Lenglart, J. Phys. Chem. Solids 30, 2409 (1969).
[24] D. Rouhani, M. Lannoo, and P. Lenglart, in *Radiation Effects in Semiconductors*, edited by J. W. Corbett and G. D. Watkins (Gordon and Breach, London, 1971), p. 15.
[25] F. Bassani, G. Iadonisi, and B. Preziosi, Phys. Rev. 186, 735 (1969).
[26] M. Jaros and S. Brand, Phys. Rev. B 14, 4494 (1976).
[27] U. Lindefelt, J. Phys. C 11, 3651 (1978).
[28] J. Bernholc and S. T. Pantelides, Phys. Rev. B 18, 1780 (1978).
[29] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. 41, 895 (1978).
[30] G. A. Baraff and M. Schlüter, Phys. Rev. Lett. 41, 892 (1978); Phys. Rev. B 19, 4965 (1979).
[31] F. Garcia-Moliner, in *Theory of Imperfect Crystalline Solids: Trieste Lectures 1970* (IAEA, Vienna, 1971).
[32] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[33] B. A. Lippman and J. Schwinger, Phys. Rev. 79, 469 (1963).

[34]J. F. Janak, A. R. Williams, and V. L. Moruzzi, Phys. Rev. B 11, 1522 (1975).
[35]In general, the defect potential extends further than the change in the charge density due to multipole fields.
[36]A systematic and practical way to construct Wannier functions for the valence bands has recently been reported by E. O. Kane and A. B. Kane [Phys. Rev. B 17, 2691 (1978)].
[37]E. O. Kane, Phys. Rev. B 13, 3478 (1976); D. J. Chadi, *ibid.* 16, 3572 (1977).
[38]R. C. Chaney, C. C. Lin, and E. E. Lafon, Phys. Rev. B 3, 459 (1971); J. Langlinais and J. Callaway, *ibid.* 5, 124 (1972); S. Ciraci and I. P. Batra, *ibid.* 15, 3254 (1977).
[39]R. Zeller and P. H. Dederichs, Phys. Rev. Lett. 42, 1713 (1979).
[40]P. O. Löwdin, J. Chem. Phys. 18, 365 (1950).
[41]J. A. Appelbaum and D. R. Hamann, Phys. Rev. B 8, 1977 (1973).
[42]M. Schlüter, J. R. Chelikowsky, S. G. Louie, and M. L. Cohen, Phys. Rev. B 12, 4200(1975).
[43]S. G. Louie (private communication).
[44]P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964); W. Kohn and L. J. Sham, *ibid.* 140, A1133 (1965).
[45]D. J. Chadi and M. L. Cohen, Phys. Rev. B 8, 5747 (1973).
[46]S. G. Louie, K. M. Ho, and M. L. Cohen, Phys. Rev. B 19, 1774 (1979).
[47]G. Gilat and L. J. Raubenheimer, Phys. Rev. 144, 390 (1966).
[48]J. F. Janak, in *Computational Methods in Band Theory*, edited by P. M. Marcus, J. F. Janak, and A. R. Williams (Plenum, New York, 1971), p. 323.
[49]J. W. Cooley and J. Bernholc (unpublished).
[50]Alternatively, the basis functions and the potential could be fitted to functions for which analytical expressions for integrals entering the calculation are available. (e.g., Gaussians (Refs. 38 and 30)). Also, the Poisson equation can be solved analytically in the Gaussian basis. The exchange-correlation potential, however, must be treated numerically and thereafter *fitted* to Gaussians. The fitting of a function in three dimensions is a rather cumbersome procedure, because of the great accuracy required. Furthermore, additional difficulties arise if the problem at hand requires a large number of Gaussians. In this case, the charge density being a bilinear function of the basis contains an impractical number of terms and requires refitting with a new set of Gaussians which places an even stronger requirement on the accuracy of the fit. A direct numerical integration scheme is much easier to program, to automate, and to test for convergence. We also believe that numerical sampling of 1000–2000 points per atom represents the potential much more accurately than the fitted function, thereby making a numerical evaluation of the matrix elements not only a simpler, but also a more accurate procedure.
[51]J. W. Cooley and J. W. Tukey, Math Comput. 19, 297 (1965).
[52]The efficiency of this procedure stems from the fact that ∇^2 is diagonal in the plane-wave representation.
[53]This fact and the results displayed in Figs. 2 and 5, which show that the total change in the charge density and vacancy potential have a predominantly single-center character, have led U. Lindefelt and S. T. Pantelides [Solid State Commun. 30, 631 (1979)] to suggest a single-center basis set consisting of harmonic-oscillator eigenfunctions as an alternative to the multicenter LCAO basis set used in the present work.
[54]For a review of tight-binding Hamiltonians, see S. T. Pantelides and J. Pollmann, J. Vac. Sci. Technol. 16, 1349 (1979).
[55]See also Ref. 37.
[56]S. T. Pantelides and W. A. Harrison, Phys. Rev. B 11 3006 (1975).
[57]J. Bernholc, N. O. Lipari, and S. T. Pantelides (unpublished).
[58]S. T. Pantelides, Solid State Commun. 30, 65 (1979).
[59]The importance of the umklapp terms was first pointed out by K. Shindo and H. Nara, J. Phys. Soc. Jpn. 40, 1640 (1976); M. Altarelli, W. Y. Hsu, and R. A. Sabatini, J. Phys. C 10, L605 (1977).
[60]S. T. Pantelides, N. O. Lipari, and J. Bernholc, Solid State Commun. 33, 1045 (1980).
[61]See also Lindefelt and Pantelides, Ref. 53.
[62]We found it convenient to orthogonalize our basis orbitals ahead of time. One could use nonorthogonal orbitals, in which case the inverse of the overlap matrix is inserted between the matrices of operators that are to be multiplied.

Solid State Communications, Vol.33, pp.1045-1049.
Pergamon Press Ltd. 1980. Printed in Great Britain.

Appendix K

THE EFFECTIVE-MASS NATURE OF DEEP-LEVEL POINT-DEFECT STATES IN SEMICONDUCTORS

S. T. Pantelides, N. O. Lipari and J. Bernholc

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA

(Received 17 December 1979 by J. Tauc)

The basic premise of Effective-Mass Theory (EMT) is that bound-state wavefunctions are constructible from Bloch functions in a small region or regions of **k** space. In contrast, deep-level wavefunctions are believed to involve Bloch functions from the entire Brillouin zone and several bands. In this paper we analyse the wavefunction of the deep vacancy level in Si obtained recently by self-consistent Green's-function calculations. We find that this wavefunction has a strong EMT character in that it is composed primarily of Bloch functions from the nearest bands and the corresponding coefficients, i.e. the envelope functions, are peaked about the band extrema. As a further check, we have used a spherical average of the self-consistent vacancy potential in the acceptor EMT equations. The resulting energy level is at E_v+0.9 eV, as compared with the Green's-function-theoretic value of E_v+0.8 eV. The resulting wavefunction, on the other hand, does not have the correct form. A check of the correction terms left out by the standard EMT equations reveals that their contributions to the energy level are large and tend to cancel one another.

Impurities which introduce bound states in the fundamental gap of a semiconductor with energy levels very near the valence- or conduction-band edge, known as shallow levels, are described very well by effective-mass theory (EMT).[1] The basic premise of that theory is that the bound-state wavefunctions can be constructed from Bloch functions in the vicinity of the nearest band edges. When first introduced, the EMT made use of a screened Coulomb potential of the form $-e^2/\epsilon r$, where ϵ is the dielectric constant, to describe the impurity potential. Later, Pantelides and Sah[2] suggested that the EMT could also be used with impurity potentials constructed from explicit pseudopotentials for individual impurity and host atoms. Such potentials have a Coulombic *tail* of the form $-e^2/\epsilon r$, but have distinctly different behavior in the so-called central cell. The EMT was found to be quite successful even for some *deep* levels, i.e., levels in the midgap region. Despite this success, however, the theory lacked fundamental justification because it was based on the *assumption* that, even for deep levels, the wavefunctions are constructible from Bloch functions in a small region or regions of the Brillouin zone at the absolute extremum of the nearest band (or few degenerate bands). This assumption was contrary to the belief that an expansion of deep-level wavefunctions in terms of Bloch functions is likely to require a large section of the Brillouin zone and several bands.[3] A few years ago, Jaros and Ross[4] adopted the latter point of view and carried out deep-level calculations by directly expanding the bound-state wavefunction in terms of Bloch functions of many bands, choosing a uniform grid of **k** points in the Brillouin zone. These calculations, as well as subsequent ones by Jaros[5] using the method of Bassani *et al.*,[1] yielded wavefunctions which, in general, were not consistent with the EMT assumption.[6] It is important to note, however, that the results obtained by Jaros and coworkers do not provide conclusive resolution of the fundamental quandary, because a low density of **k** points was used and the calculations were not carried to full convergence.[7]

We have recently developed[8] a computational procedure which permits the unambiguous resolution of this question. The procedure, based on a Green's-function formalism, produces self-consistent and convergent solutions and thus yields accurate wavefunctions as well as bound-state energies. We have so far reported results on the electronic structure of the vacancy in Si.[8] (Virtually identical results obtained by a similar method have been reported independently by Baraff and Schlüter[9]). In these calculations, the wavefunctions are determined in terms of a linear-combination-of-atomic-orbitals (LCAO) basis set. In Ref. 8, we displayed the bound-state wavefunction of the vacancy in Si in terms of a contour plot in real space. Once the wavefunction is known, however, it is straightforward to project it on Bloch functions and calculate the corresponding ex-

pansion coefficients. We have therefore calculated the quantities $F_n(\mathbf{k})$ defined by

$$\psi(\mathbf{r}) = \sum_{n\mathbf{k}} F_n(\mathbf{k})\psi_{n\mathbf{k}}(\mathbf{r}) \qquad (1)$$

by expressing them as

$$F_n(\mathbf{k}) = \langle\psi_{n\mathbf{k}}|\psi\rangle \ . \qquad (2)$$

Using standard Green's-function expressions for ψ (see, e.g., the review article by Pantelides, Ref. 1), we get the simple result

$$F_n(\mathbf{k}) = \langle\psi_{n\mathbf{k}}|U|\psi\rangle/(E_B - E_{n\mathbf{k}}) \qquad (3)$$

where $E_{n\mathbf{k}}$ are the band energies, E_B is the bound-state energy, and U is the perturbation potential of the impurity or defect.

The results of this calculation are shown in Fig. 1. We plot $|F_n(\mathbf{k})|^2$ for the bands corresponding to n = 2, 3, 4, 5, and 6. The contributions of the other bands are several orders of magnitude smaller. It is clear that, even though the unrelaxed vacancy corresponds to one of the strongest possible point-defect perturbations and results in one of the most tightly-bound states in the gap,[10] the bound-state wavefunction has a strong EMT character in k-space.[11] Since the vacancy introduces a perturbation which is repulsive to electrons, it behaves like an acceptor impurity with $\Delta z = 4$ (the neutral vacancy may be thought of as binding four holes). The envelope functions $F_n(\mathbf{k})$ are thus peaked at $\mathbf{k} = 0$ for the three degenerate valence bands. Note also the small contribution from the lowest two conduction bands in the vicinity of the X point and, in particular, the peaking of $F_5(\mathbf{k})$ at the minimum of the lowest conduction band.

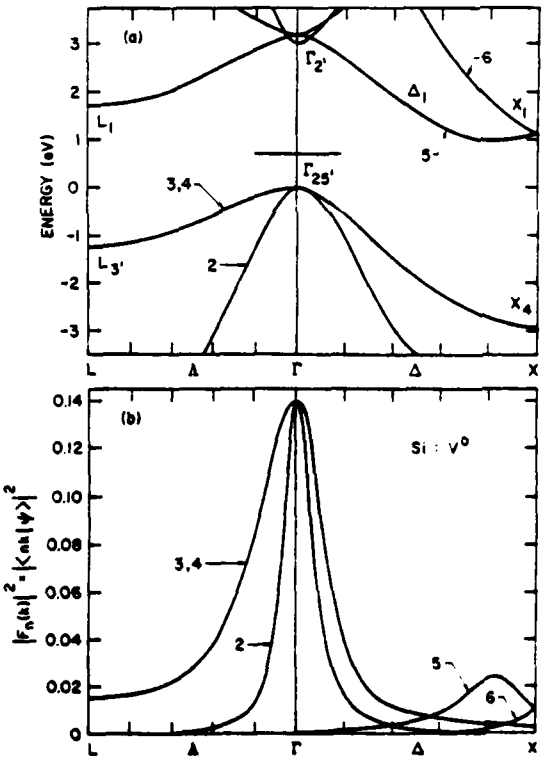


Fig. 1. (a) The energy bands of Si along two symmetry directions and the vacancy level in the gap. (b) The k-space decomposition of the bound-state wavefunction of the vacancy in Si, calculated by the Green's-function method (Ref. 8).

As we remarked earlier, the unrelaxed vacancy corresponds to one of the strongest point perturbations. Perhaps the strongest point perturbation is one that couples each Bloch function in a set of bands with all other Bloch functions in that set by the same matrix element V. In that case, Eq. (3) reduces to the simple result[16]

$$F_n(\mathbf{k}) = C/(E_B - E_{n\mathbf{k}}) \qquad (4)$$

where C is a constant independent of n and k. We have plotted these quantities in Fig. 2 and observed that the overall EMT character remains unchanged.

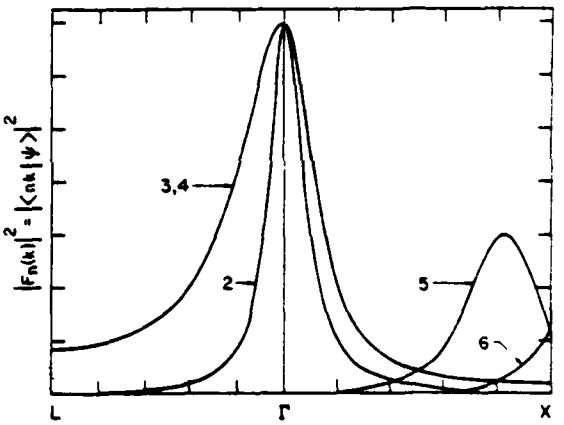


Fig. 2. The k-space decomposition of the bound-state wavefunction arising from a model perturbation that couples all Bloch states in a set of bands by the same matrix element (see text).

The plots shown in Figs. 1 and 2 should be interpreted with caution, because they show the $F_n(\mathbf{k})$ only along two symmetry lines. For an assessment of the total contribution of each band, sums over the Brillouin zone were carried out. The results are shown in Table I. These results confirm the conclusion that a state arising from a repulsive potential is constructed primarily from the top valence bands.

The strong EMT character of the bound-state wavefunction suggests that one might, in fact, be able to use the EMT equations to determine the bound-state energy. In order to check this possibility, we have used a spherical vacancy potential taken to be the negative of a bulk self-consistent atomic

TABLE I

The per-cent contribution of each Si band to the vacancy bound-state wavefunction. The quantity c_n is defined by $\sum_{\mathbf{k}} |F_n(\mathbf{k})|^2$.

n	c_n [%]
1	0.4
2	4.8
3	23.7
4	51.1
5	10.9
6	4.1
7	2.2
8	1.4
9	0.6
10	0.3

pseudopotential (Fig. 3). This potential, when used in the Green's-function equations yields a bound state at 0.8 eV, compared with the fully self-consistent value of 0.7 eV (we avoided using the fully self-consistent vacancy potential because it has nonspherical terms that would unnecessarily complicate the EMT calculation). We have used the acceptor EMT scheme developed by Lipari and Baldereschi,[12] incorporating the full anisotropy at the valence-band top and allowing complete variational freedom for the envelope functions $F_n(\mathbf{r})$ [which are Fourier transforms of the $F_n(\mathbf{k})$]. (Note that the vacancy potential we used has no Coulomb tail. It is a totally central-cell potential.) The result for the bound-state energy was astonishingly good. We obtained a level at 0.9 eV above the valence band edge, which is to be compared with the corresponding Green's-function-theoretic value of 0.8 eV. The corresponding wavefunction, however, did not have the correct structure. It is compared with the wavefunction obtained from the Green's-function calculation in Fig. 4. In the top panel, we show the Bloch functions at the top of the valence bands in order to illustrate that the EMT bound-state wavefunction retains the overall Bloch-function structure (since it is in fact a product of envelope functions and Bloch functions).

The above results led us to investigate the size of the terms that are left out by the standard acceptor EMT.[1,12] These terms are: (i) the so-called Umklapp terms in the potential matrix elements;[13-15] (ii) the k^4 and higher-order terms in the expansion of the valence bands around $\mathbf{k}=0$ which enter the EMT kinetic-energy matrix elements; and (iii) the conduction-band contributions.

The Umklapp terms arise in the potential matrix elements when one does not make a plane-wave approximation to the Bloch functions, but, instead, keeps all the terms in the expansion of reciprocal-lattice-vector plane waves. We have estimated the Umklapp contributions by perturbation theory by evaluating the expression

$$\Delta E = \sum_{n\mathbf{k}} \sum_{n'\mathbf{k}'} F_n(\mathbf{k})[\langle \psi_{n\mathbf{k}} | U | \psi_{n'\mathbf{k}'} \rangle - \langle \mathbf{k} | U | \mathbf{k}' \rangle \delta_{nn'}] F_{n'}(\mathbf{k}') \quad (5)$$

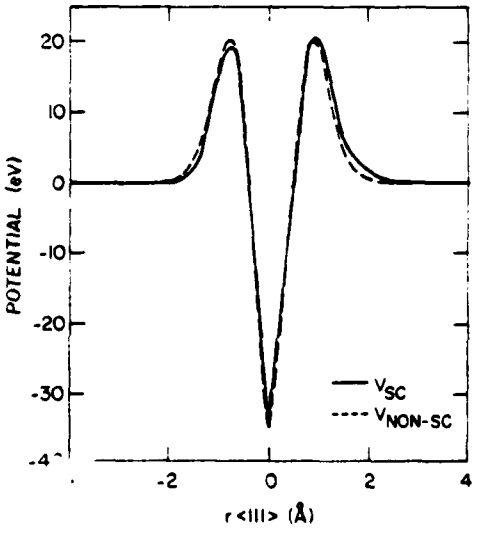


Fig. 3. The spherical vacancy potential used in the EMT calculations compared with the full self-consistent potential along the (111) axis.

where $|\mathbf{k}\rangle$ stands for a plane wave. The sum over n is over the top three valence bands and $F_n(\mathbf{k})$ are the envelope functions obtained from the EMT calculation described above. We found that the Umklapp correction is very large, of order 2 eV. This result indicates that the success of the EMT calculation described above for the bound-state energy is a consequence of approximate cancellations of large correction terms. Since the k^4 kinetic-energy corrections would, like the Umklapp terms, make the level deeper, it appears that the conduction-band contributions are the ones that would compensate the large Umklapp terms. The conduction-band contributions, which in this case make up about 20% of the wavefunction (Table I), can in principle be incorporated in an EMT calculation (see, e.g., the multiband and multivalley formulation of Ref. 15), but the task is tedious and beyond the scope of the present paper. Some insight into their importance can be gained, however, by comparing the bound-state wavefunction obtained by the EMT calculation with that obtained by the Green's-function calculation (Fig. 4). It is

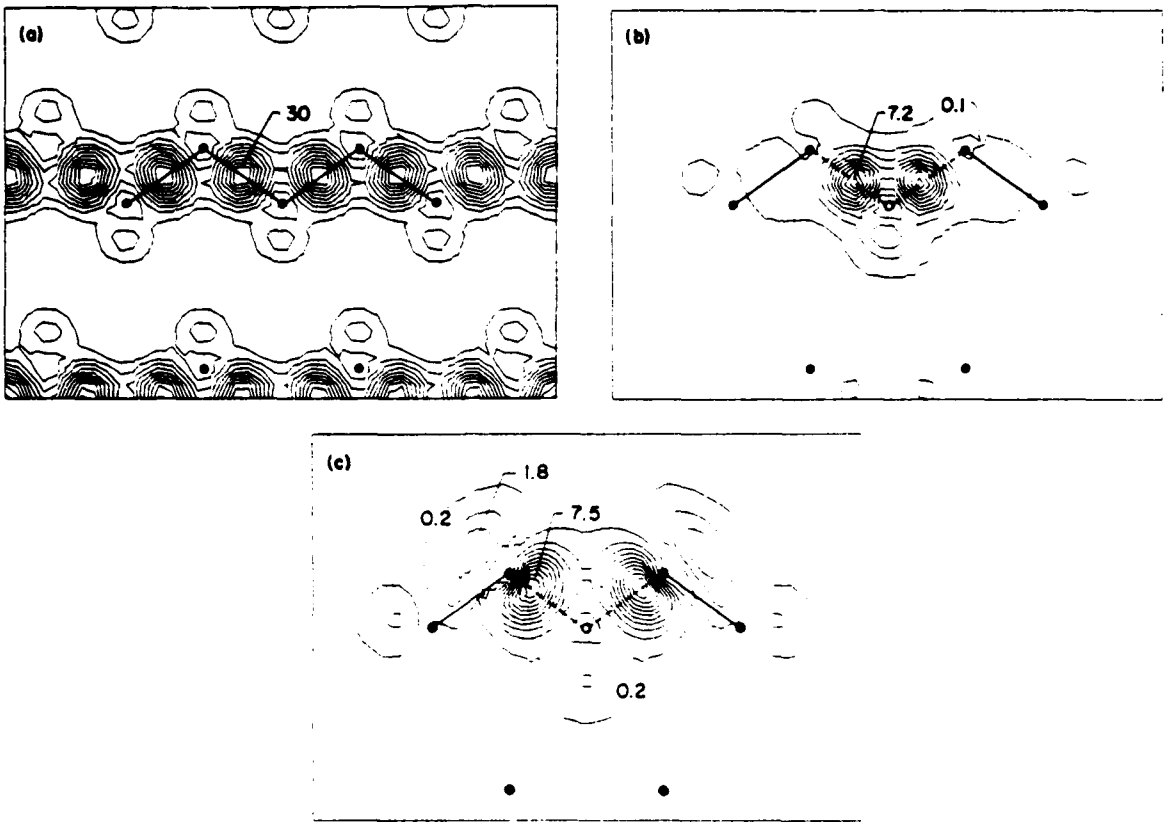


Fig. 4. Contour plots on the (110) plane of (a) Bloch functions at the top of the valence bands (sum of the three degenerate Bloch functions at $\mathbf{k} = 0$); (b) the EMT bound-state wavefunction; and (c) the Green's-function-theoretic bound-state wavefunction (from *Ref. 8*). *The units are electrons per primitive* unit cell. [Note that in (a) the triply degenerate state contains six electrons, whereas the triply degenerate state in (b) and in (c) contains only two electrons].

clear that, although the Green's-function-theoretic wavefunction has a dangling-hybrid-like character, the EMT wavefunction does not. Instead, it looks like a bonding function with maximum amplitude in the region where the vacancy potential (Fig. 3) is large. Admixture of conduction-band Bloch functions, which are antibonding in character, will, however, shift the amplitude toward the nearest neighbors.[17] Such a shift can be expected to reduce the potential matrix element substantially. These qualitative arguments, however, do not explain why the conduction-band contributions cancel the Umklapp terms of the valence-band contributions so effectively.

In conclusion, the studies reported in the present paper have established that: (a) The wavefunctions of deep point-defect levels are strongly effective-mass-like in that the k-space envelopes are strongly peaked at the nearest band extrema. (b) For deep levels, the EMT equations without Umklapp terms can provide energy levels that compare well with more sophisticated calculations, even though the corresponding wavefunctions disagree substantially. The good energy levels result from cancellations of large corrections. The surprising effectiveness of these cancellations in the case of the vacancy is not, however, completely understood at present and cannot, therefore, be generalized. (c) EMT equations with Umklapp terms may yet prove to be quantitative even for very deep levels, but both valence and conduction bands would have to be incorporated in the calculation.

Acknowledgement: This work was supported in part by the Air Force Office of Scientific Research under Contract No. F49620-79-C-0077.

REFERENCES

1. W. Kohn, Solid State Phys. **5**, 257 (1957). For recent reviews see F. Bassani, G. Iadonisi and B. Preziosi, Rep. Prog. Phys. **37**, 1099 (1974), and S. T. Pantelides, Rev. Mod. Phys. **50**, 797 (1978).

2. S. T. Pantelides and C. T. Sah, Solid State Comm. **11**, 1713 (1972); Phys. Rev. B **10**, 621 (1974); *ibid.*, p. 638; see also S. T. Pantelides, in *Festkörperprobleme*, ed. by H. J. Queisser (Pergamon/Vieweg, Braunschweig, 1975), vol. XV, p. 149.

3. See, e.g., Bassani *et al.*, Ref. 1.

4. M. Jaros and S. F. Ross, J. Phys. C **6**, 3451 (1973); Proc. of the 12th Intern. Conf. on the Phys. of Semiconductors, ed. by M. H. Pilkuhn (Teubner, Stuttgart, 1974), p. 401.

5. See, e.g., M. Jaros, J. Phys. C **8**, 2455 (1975).

6. See, e.g., Table 3 of Ref. 4(b) and Table 1 of Ref. 5. These authors find that the contributions of the lowest conduction band have a maximum at Γ and a secondary maximum (smaller by a factor 2) at X whereas an EMT picture would suggest the opposite behavior, i.e., a strong maximum at X and a weak secondary maximum at Γ because the band extremum at Γ is higher in energy and has a very small effective mass. Similarly, Ref. 5 finds the top-valence-band contribution to peak away from $k=0$, contrary to EMT expectations. Also against EMT expectations is the result (Ref. 5) that valence band number 2 (i.e., the third band below the top valence band, known as the light hole band because of its small effective mass) contributes strongly whereas the heavy-hole band contributes very weakly.

7. In the early work by Jaros and coworkers the convergence of the calculations was not tested thoroughly. More recently, additional tests were carried out. J. Jaros, C. O. Rodriguez, and S. Brand [Phys. Rev. B **19**, 3137 (1979)] have concluded that the precision with which they are prepared to execute their calculations does not seem adequate to establish precise positions of the bound states in the forbidden gap. In general, variational wavefunctions are less accurate than the corresponding energy levels. In addition, the basis functions used by Jaros and coworkers are divergent at the origin [see discussion by U. Lindefelt and S. T. Pantelides, Solid State Comm. **30**, 631 (1979)] and therefore do not provide an accurate representation of the wavefunction.

8. J. Bernholc, N. O. Lipari and S. T. Pantelides, Phys. Rev. Lett. **41**, 895 (1978).

9. G. A. Baraff and M. Schlüter, Phys. Rev. Lett. **41**, 892 (1978).

10. The unrelaxed vacancy corresponds to one of the strongest point-defect perturbations because substitutional impurities in general reduce the strength of the perturbation. Lattice relaxation around the vacancy, however, may weaken the total perturbation.

11. "EMT character" *does not* in this case imply hydrogenic character. As stated in the beginning of this paper, the basic EMT assumption is that the bound-state wavefunction is constructible from Bloch functions from mainly around the extrema of the nearest bands. Use of a hydrogenic potential ($-e^2/\epsilon r$) is an *additional* assumption that corresponds to a special application of the EMT equations.

12. N. O. Lipari and A. Baldereschi, Phys. Rev. Lett. **25**, 1660 (1970); Solid State Commun. **25**, 665 (1978).

13. K. Shindo and H. Nara, J. Phys. Soc. Japan **40**, 1640 (1976).

14. M. Altarelli, W. Y. Hsu and R. A. Sabatini, J. Phys. C **10**, L605 (1977).

15. S. T. Pantelides, Solid State Comm. **30**, 65 (1979).

16. An example of such a case is the tight-binding description of the unrelaxed vacancy given by J. Bernholc and S. T. Pantelides, Phys. Rev. B **18**, 1780 (1978).

17. If h_1 is an sp^3 hybrid on the atom to be removed and h_2 is an sp^3 hybrid on a nearest neighbor, then $b=h_1+h_2$ is a bonding orbital and $a=h_1-h_2$ is an antibonding orbital. Consequently, admixing antibonding orbitals in a bonding-like wavefunction will tend to give it a dangling-hybrid-like structure.

VOLUME 43, NUMBER 18 PHY . . . . .S 29 OCTOBER 1979

Electronic Structure of the Jahn-Teller Distorted Vacancy in Silicon

N. O. Lipari, J. Bernholc, and S. T. Pantelides
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 20 July 1979)

This Letter reports self-consistent-field calculations of the electronic structure of the Jahn-Teller distorted vacancy in Si. With use of the tetragonal atomic displacements estimated by Watkins, it is found that the Jahn-Teller splitting of the sixfold-degenerate bound state in the gap is of the order of 0.5 eV. This, together with small breathing-mode displacements, results in a fully occupied doublet in the lower part of the band gap, in agreement with experimental observations.

The introduction of point defects such as vacancies and many chemical impurities in semiconductors gives rise to substantial rearrangements of the electron density and the atomic positions which are not well described by effective-mass and linear-response theories. The resulting localized states possess energy levels that lie deep in the forbidden energy gap and play an important role in determining many device properties through their influence on carrier lifetimes and impurity diffusion. Recently, the authors[1] and Baraff and Schlüter[2] independently reported the development of self-consistent Green's-function techniques which, for the first time, gave a detailed description of the electronic structure of the unrelaxed vacancy in Si at the same level of sophistication and accuracy characteristic of state-of-the-art electronic-structure calculations for bulk semiconductors, surfaces, and interfaces. However, that work did not take into account lattice distortions in the immediate vicinity of the vacancy. Such distortions, which are a manifestation of the Jahn-Teller effect, are known experimentally[3] to have a significant effect on the properties of the vacancy. The purpose of this paper is to provide a detailed theoretical picture of the consequences of Jahn-Teller distortions in terms of calculations which retain the level of accuracy achieved in the case of the undistorted vacancy.[1] We show that small, symmetry-breaking displacements of the nearest neighbors (of order a few tenths of an angstrom, as estimated by simple force models) have large effects on the electronic structure of the vacancy. In particular, they split the sixfold-degenerate bound state in the gap into a fully occupied doublet and an empty quadruplet. The splitting is large (of the order of half the band gap), so that the energy level of the occupied doublet is in the lower part of the band gap, in agreement with experimental observations.

From an analysis of electron-spin-resonance (ESR) data, Watkins[3] has concluded that the preferred distortion is tetragonal in nature, lowering the point symmetry from T_d to D_{2d}. Watkins also estimated the magnitudes of the displacements of the nearest neighbors with use of a simple force model. In addition to the symmetry-breaking displacements, the nearest neighbors can also move toward or away from the vacant site in the symmetric, so-called breathing mode, but no estimates of the corresponding displacements have been made.

Our calculations permit us to study the effects of symmetry-lowering and breathing distortions, independently, as well as simultaneously. The calculations were carried out using the method of Ref. 1,[4] the only difference being that the ionic contribution to the perturbation pseudopotential is now given by

$$U_i(\vec{r}) = -v_i(\vec{r}) - \sum_{\vec{R}} v_i(\vec{r}-\vec{R}) - \sum_{\vec{R}'} v_i(\vec{r}-\vec{R}'). \quad (1)$$

Here $v_i(\vec{r})$ is a Si^{4+} ionic pseudopotential, $\vec{R}$ are the undistorted-lattice nearest-neighbor positions, and $\vec{R}'$ are the new nearest-neighbor positions. [In the case of the unrelaxed vacancy,[1] $U_i(\vec{r})$ contained only the first term in Eq. (1).] For each choice of the $\vec{R}'$, the potential arising from the change in the valence charge density is calculated self-consistently.

We find that a tetragonal distortion of the magnitude estimated by Watkins splits the sixfold-degenerate T_2 level, which in the undistorted case lies at $E_v + 0.7$ eV (where E_v is the valence-band edge), into a doublet B_2 at $E_v + 0.3$ eV and a quadruplet E at $E_v + 0.8$ eV. The doublet contains two electrons which maintain the neutrality of the defect and the quadruplet is empty. This result demonstrates explicitly that small symmetry-breaking displacements of the nearest neighbors of the vacancy can produce large level splittings.[5] Indeed, the calculated splitting, 0.5 eV, is of the order of half the band gap. Allowing for the pos-

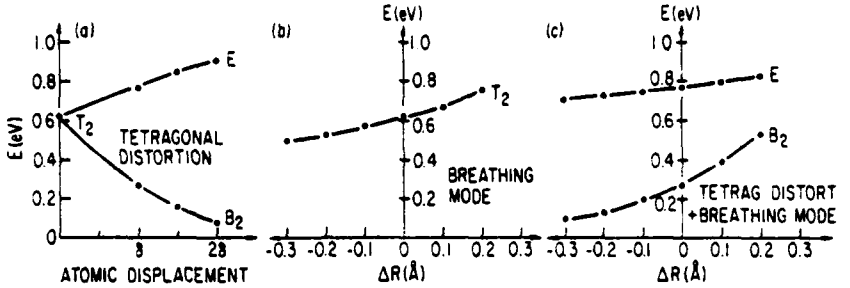


FIG. 1. (a) Jahn-Teller splitting of the T_2 vacancy level in the gap for various values of tetragonal displacements of the nearest neighbors. δ is equal to the value estimated by Watkins (Ref. 3). (b) The effect of breathing-mode displacements of the vacancy nearest neighbors on the T_2 vacancy level in the gap. Negative values correspond to inward displacements. (c) The combined effect of tetragonal displacements (as estimated by Watkins, Ref. 3) and various breathing-mode displacements of the vacancy nearest neighbors on the T_2 vacancy level in the gap. Negative values correspond to inward displacements.

sibility that Watkins's values of the displacements represent underestimates,[3] we repeated the calculation with larger displacements. The results are shown in Fig. 1(a).

In addition to symmetric-breaking displacements, we have investigated the effect of breathing-mode displacements of the nearest neighbors. The results, for several values of the atomic displacements, are shown in Fig. 1(b). As expected, no splitting occurs. The T_2 level simply moves up or down depending on whether the atoms move away from or toward the vacant site. We have also carried out calculations in which tetragonal distortions are combined with breathing-mode displacements. One set of results is shown in Fig. 1(c). In all cases, we find that inward breathing displacements lower the energies of both the doublet and the quadruplet, whereas outward breathing displacements raise both levels. Overall, the calculations indicate that a combination of tetragonal and breathing displacements can indeed place the energy level of the fully occupied doublet in the lower part of the gap, as found experimentally.[6] As Fig. 1(c) shows, this can be accomplished by combining Watkin's estimates for the tetragonal displacements with small inward breathing-mode displacements, or by combining larger tetragonal displacements with small outward breathing-mode displacements. The latter possibility is consistent with arguments suggesting that outward displacements, which strengthen the backbonds, are most likely to occur.[7] If that is indeed the case, our calculations indicate that the preferred breathing-mode displacements actually raise the energy level of the fully occupied doublet. This rather surprising behavior is actually consistent with the fact that the bound states in the gap have a predominantly dangling-bond character (Fig. 2). Detailed understanding of this

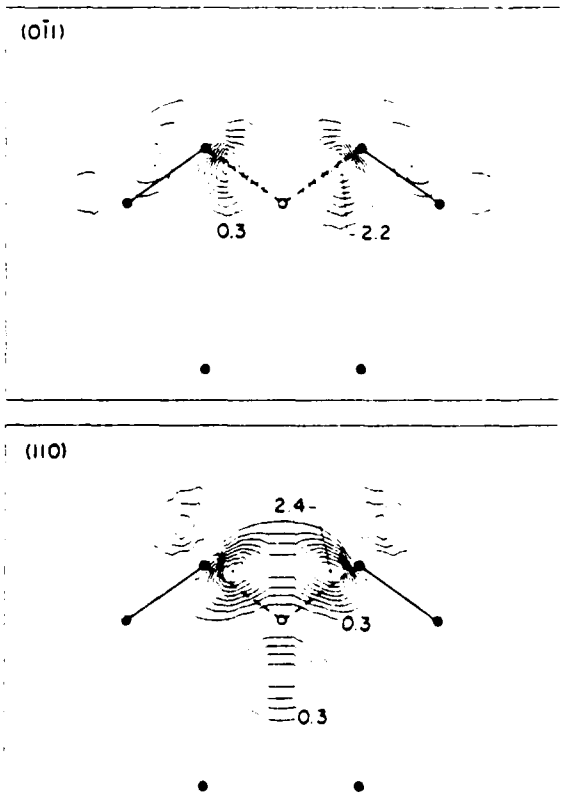


FIG. 2. The wave function of the B_2 bound state of the tertagonally distorted vacancy in Si in two inequivalent (110) planes.

behavior can be obtained in terms of a simple tight-binding model,[8] according to which both the T_2 bound state and the A_1 resonance at approximately $E_v - 0.7$ eV (Ref. 1) are basically linear combinations of the sp^3 dangling hybrids on the nearest neighbors, and the T_2-A_1 splitting is determined by the interaction V between dangling hybrids. When the atoms move in or out, two changes occur: First, the hybrid energy changes because its composition is no longer pure sp.[3] This dehybridization energy E_d can be estimated,[9] and, in fact, is found to lower the dangling-hybrid energy when the nearest neighbors move toward the vacancy. The second change occurs in the value of V. When the atoms move toward the vacancy, V is enhanced and the T_2-A_1 splitting increases. Thus, we conclude[10] that inward motion would result in a lowering of the A_1 resonance and an increase in the T_2-A_1 splitting. The T_2 level could move either way depending on the ratio of E_d and V. The opposite effects would be produced by outward motion. These simple predictions are confirmed by our self-consistent results (Fig. 3).

The present calculations also yield detailed information about individual wave functions and charge-density changes. In Figs. 2 and 4 we show some results in the case of a pure tetragonal distortion with use of Watkin's estimates of the atomic displacements. In Fig. 2 we show contour plots of the wave function of the occupied B_2 bound state in the gap. The two plots correspond to two inequivalent (110) planes, demonstrating a strong anisotropy of the wave function induced by the distortion. In Fig. 4 we show the total *change* in the charge density. The two plots correspond to the same two inequivalent (110) planes used in Fig. 2. We again observe a substantial amount of anisotropy caused by the tetragonal distortion. We also note that the change now extends beyond the cavity defined by the nearest neighbors. The nature of these charge-density changes depends on the type of assumed reconstruction and, in general, can be understood in terms of simple physical pictures. A detailed discussion will be given

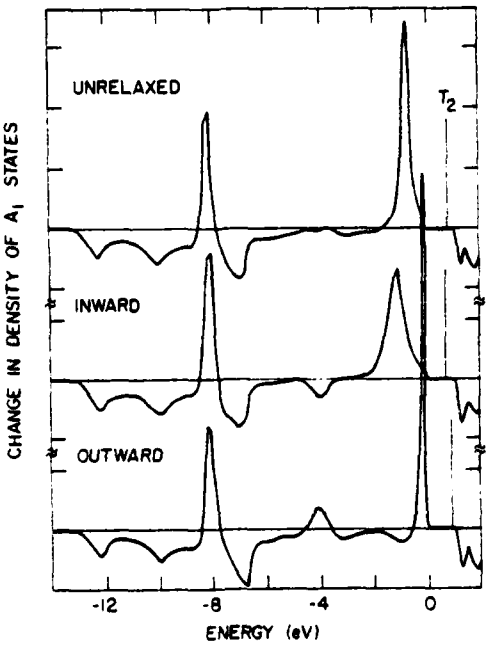


FIG. 3. The changes in the density of A_1 states within the band continua and the position of the T_2 bound state in the gap for the unrelaxed vacancy in Si, and for inward and outward breathing-mode displacement of the nearest neighbors.

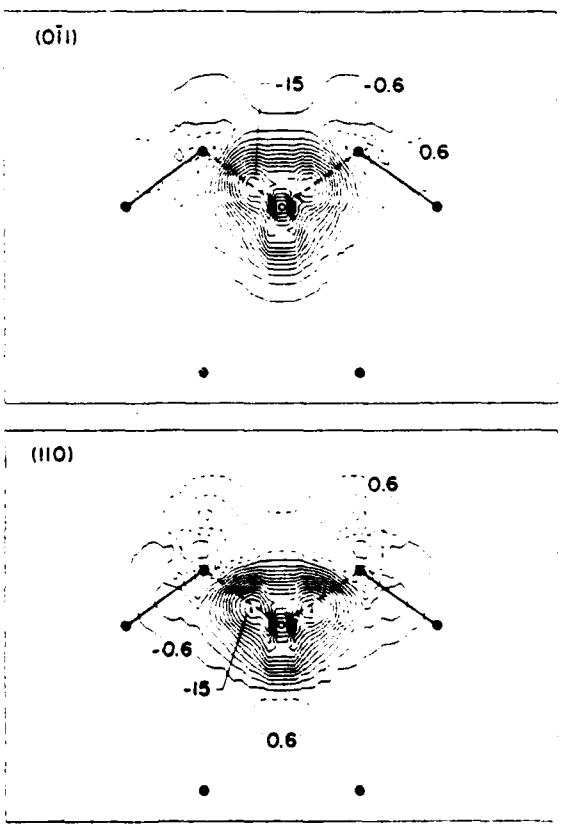


FIG. 4. The total change in the charge density in the case of the tetragonally distorted vacancy in Si in the same two inequivalent (110) planes as in Fig. 2. Solid lines correspond to removal of charge and dashed lines to addition of charge.

elsewhere.

In conclusion, we have used self-consistent-field calculations to study the effect of both symmetry-lowering and symmetry-preserving lattice distortions on the electronic structure of an isolated neutral vacancy in Si. We find that distortions of reasonable magnitude are very important and bring the bound-state energy to within the range of experimental measurements.

We would like to express our appreciation to G. D. Watkins for his interest in this work and many valuable discussions. This work has been supported in part by the U. S. Air Force Office of Scientific Research under Contract No. F49620-77-C-0005.

---

[1]J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. 41, 895 (1978).

[2]G. A. Baraff and M. Schlüter, Phys. Rev. Lett. 41, 892 (1978).

[3]G. D. Watkins, in *Lattice Defects in Semiconductors —1974*, The Institute of Physics Conference Proceedings No. 23, edited by F. A. Huntley (The Institute of Physics, Bristol and London, 1975), p. 1.

[4]The calculations were carried out using the ten-orbitals-per-atom basis of Ref. 1 and repeated with additional orbitals at sites 0.5 A from the ideal nearest-neighbor positions in order to ensure convergence.

[5]Similar results have been obtained by the authors of Ref. 2 (private communication). Baraff, Kane, and Schlüter further adopted a phenomenological model for elastic forces and estimated the atomic displacements for various charge states of the vacancy.

[6]Experimentally, G. D. Watkins, J. R. Troxell, and A. P. Chatterjee [In, *Defects and Radiation Effects in Semiconductors—1978*, The Institute of Physics Conference Proceedings No. 46, edited by J. H. Albany, The Institute of Physics, Bristol and London, 1979), p. 16] identify a level at $E_v - 0.13$ eV, which they assign to either the $V^- \rightarrow V^0$ or the $V^0 \rightarrow V^+$ transition (see their Fig. 5). The energy level of the neutral vacancy is thus deduced to lie between $E_v + 0.04$ and $E_v - 0.04$ eV.

[7]The main argument is based on Pauling's bond-order/bond-length relation. For a summary of all arguments, see J. A. Van Vechten, Phys. Rev. 10, 1482 (1974). Semiempirical cluster calculations [e.g., R. P. Messmer and G. D. Watkins, Phys. Rev. B 7, 2568 (1973); K. L. Yip, Phys. Status Solidi (b) 66, 619 (1974)] have arrived at the same conclusion. Our calculations show that the backbond charge decreases when the nearest neighbors are moved toward the vacant site and increases in the opposite case. This result is consistent with a similar finding for surfaces [see, e.g., J. A. Appelbaum and D. R. Hamann, Phys. Rev. Lett. 31, 106 (1973), and Phys. Rev. B 8, 1777 (1973)]. See also the discussion by S. G. Louie, M. Schlüter, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. B 13, 1654 (1976).

[8]J. Bernholc and S. T. Pantelides, Phys. Rev. B 18, 1780 (1978). See also J. Bernholc, N. O. Lipari, and S. T. Pantelides, to be published, and compare with the defect-molecule model, first introduced by C. A. Coulson and M. J. Kearsley, Proc. Roy. Soc. London, Ser. A 241, 433 (1957).

[9]See, e.g., W. A. Harrison, Surf. Sci. 55, 1 (1976).

[10]A more detailed discussion of these arguments will be given elsewhere.

[11]See Messmer and Watkins, Ref. 7.

Appendix M

THE ELECTRONIC STRUCTURE OF DEEP SP-BONDED IMPURITIES IN SEMICONDUCTORS

J. Bernholc, S. T. Pantelides, N. O. Lipari, and A. Baldereschi*

IBM Thomas J. Watson Research Center

Yorktown Heights, NY 10598

This Letter reports self-consistent calculations of the electronic structure of a series of deep impurities in Si. For the first time, these calculations provide a detailed description of gap states as well as the resonances and antiresonances within the band continua with the same accuracy as that of bulk-crystal calculations. The analysis of charge densities and wavefunctions in terms of simple models provides an understanding of the chemical bonding, the degree of localization, and the relevance of ideas based on effective-mass theory.

Impurities in semiconductors usually introduce bound states in the fundamental gap and through them control transport and other properties important to the fabrication and performance of electronic devices. Effective-mass theory (EMT)[1-3] has been quite successful in describing such states if the impurity potential is dominated by a screened Coulombic tail. Most such impurities have energy levels very near a band edge and are thus known as shallow. Many other impurities, however, introduce strong central-cell potentials which result in bound states in the midgap region, and are thus referred to as deep. The EMT has offered some insight into some of these states,[2,3] but, usually, alternative techniques are found to be more appropriate.[2] Until now, however, all such techniques have had to make compromising approximations and hence resulted in incomplete or quasiquantitative descriptions.[4] For example, electronic redistribution in the vicinity of the impurity is often either neglected or treated by linear-response theory.

In this Letter we report detailed calculations of the electronic structure of a series of deep impurities, carried out with the same rigor and accuracy that is currently possible for perfect crystals. The complete spectrum of localized states is calculated, including gap states as well as resonances and antiresonances within the valence-band continuum. The results are analysed in terms of simple physical models which allow us to draw conclusions about chemical bonding, localization, and the applicability of effective-mass ideas.

The series of impurities we study in this Letter are nominal acceptors in the sense that they are substitutional and have a smaller chemical valence than the host. If we define $\Delta z = z_{imp} - z_{host}$, where z is the chemical valence (e.g., 4 for Si), $\Delta z = -1$ corresponds to the shallow acceptors (e.g., Si:Al). Impurities with $\Delta z = -2$ are often referred to as double acceptors since they

may have two ionization states (He analog). One can then proceed to $\Delta z=-3$, the nominal triple acceptors, and, finally, to $\Delta z=-4$, which is the vacancy. In this paper, we will describe substitutional[5] H and Zn in Si ($\Delta z=-3$ and -2, respectively), which will allow us to connect with the vacancy results[6] ($\Delta z=-4$) and also with the shallow, EMT limit ($\Delta z=-1$). Since these impurities, like Si, have no d valence electrons, we will be able to investigate and elucidate the restructuring of host sp bonds by sp-bonded impurities. Transition-metal impurities with d valence electrons constitute a special class and must be treated separately.

The calculations were carried out using the Green's-function method reported in Ref. 6. The only difference is that the bare defect potential is now the difference between the impurity-ion and the host-ion pseudopotentials,[7] instead of just the negative of a host-ion pseudopotential used in the case of the vacancy.[6] The valence charge density is calculated self-consistently within the Kohn-Sham local-density one-electron scheme.

The results reveal that the important changes occur for states of A_1 (s-like) and T_2 (p-like) symmetry. In particular, as in the case of the vacancy (Si:V),[6] Si:H and Si:Zn have a T_2 bound state in the gap.[8] Their energy levels and wavefunctions are shown in Fig. 1. This figure indicates that the wavefunction has a canonical shape for the entire series,[9] getting only slightly more compact as the level moves deeper in the gap. In all three cases, it consists mainly of the p-like combination of the neighbor sp^3 hybrids. The origin of this effect can be traced to the fact that H and Zn atoms, just like the vacancy, cannot support p-like valence electrons in the atomic cell. (The basis for this conclusion will become clearer below).

In addition to the bound state in the gap, the impurity potential also modifies the band continua. Consistent with the discussion above, we find that the changes in the density of T_2 states induced by H and Zn are virtually identical with those induced by the isolated vacancy.[6] The changes in the density of A_1 states induced by H and Zn, on the other hand, differ dramatically from those induced by the vacancy in that the prominent A_1 vacancy resonance at E_v-0.7 eV becomes very broad and diffuse (Fig. 2). This contrasting behavior arises because the vacancy resonance lies at an energy where the crystal has virtually no Bloch states with s-like amplitude in the atomic cell, whereas the impurity resonance is pulled at lower energies where such states exist. The charge densities of these resonances, shown in Fig. 3, clarify the picture further. Whereas the vacancy resonance is the s-like combination of the dangling hybrids,[6] just like the T_2 bound state is the p-like combination of the same hybrids, the impurity resonance has a nearly spherical s-like charge around the nucleus. Thus, when an H or Zn atom is inserted in a vacancy, its s orbital bonds with the s-like combination of the dangling hybrids, allowing the s-like propagating states to exist almost undisturbed. Equivalently, our results show that when H or Zn replace a Si atom, the crystalline s-like charge in the atomic cell is only slightly disturbed. The origin of this behavior can again be traced to the fact that H and Zn atoms, this time *unlike* the vacancy, do support s-like charge in the atomic cell.

The overall conclusion of the above analysis is that the s-like parts of the bonds broken by the removal of a Si atom are reestablished when a group-I or group-II impurity is substituted. In contrast, the p-like parts of the broken bonds remain unsaturated even as additional electrons are put into the

p-like state in the gap. This result reflects the fundamental ability of H and Zn atoms to support s-like but not p-like valence electrons in the atomic cell.

It would be desirable to compare our theoretical results with experimental data. Unfortunately, no unambiguous data are available. Experimental evidence has been mounting lately that many centers which were thought to be simple substitutional impurities are in fact complexes. The case of Au in Si, which, if substitutional, would be a $\Delta z=-3$ impurity, is the most striking example.[2] The data on Zn-doped Si, however, may correspond to substitutional Zn. Two charge states have been observed with ionization energies of 0.3 and 0.6 eV, respectively.[10] We have, therefore, repeated our calculations for Zn^- and Zn^{--}. We also determined the first and second ionization energies using the transition-state approximation.[11] Our results are 0.14 and 0.28 eV. For a more precise determination of these ionization energies, other effects which were left out for the purposes of the present study would have to be included. Such effects include lattice relaxation,[12] a self-consistent account of the Zn d electrons, spin-orbit coupling, and a more accurate treatment[13] of the Coulombic tail of the potential for charged states. None of these effects would change the general results discussed in this paper, but each could contribute corrections to the ionization energies of order 0.1 eV.

Finally, we compare the present analysis with alternative theoretical approaches. One such approach is to use a tight-binding Hamiltonian as done for the vacancy by Bernholc and Pantelides[14] and for deep impurities by Hjalmarson et al.[4] We have carried out such calculations (Hjalmarson et al.[4] reported calculations only for A_1 gap states) and find that they reproduce the state-density changes in the valence-band continuum remarkably well, as is true for Si:V,[6] but the bound-state energy level is quite sensitive to the choice

of tight-binding matrix elements for both the crystal Hamiltonian and the impurity potential. The tight-binding method, on the other hand, does not yield unambiguous information on charge-density changes.[15]

Hjalmarson et al.[4] also suggested that the important physics of deep-impurity states can be understood in terms of a "defect-molecule" model that focuses on the impurity atom and its four neighbors, neglecting the coupling with the rest of the lattice. Applying those ideas to the impurities at hand, we find that the defect molecule merely suggests that both the s-like and p-like crystal charge will be displaced toward the neighbors, as indeed is the case. It fails, however, to recognize that the effect is almost negligible for the s-like charge, but constitutes almost total exclusion for the p-like charge, as discussed above. The defect-molecule model also misses entirely the resonances at -8 eV, which are band-structure effects.[6]

ACKNOWLEDGMENTS: This work was supported in part by the Air Force Office of Scientific Research under Contract No. F49620-79-C-0077 and the Office of Naval Research under Contract No. N00014-80-C-0679.

REFERENCES

* IBM summer faculty 1978. Permanent address: Ecole Polytechnique Federale, Lausanne, Switzerland.

1. W. Kohn, Solid State Phys. **5**, 257 (1957); for more recent reviews see Refs. 2 and 3.

2. S. T. Pantelides, Rev. Mod. Phys. **50**, 797 (1978).

3. M. Altarelli, A. Baldereschi, and N. O. Lipari, Solid State Phys., to be published.

4. For a review of approaches used before 1978, see Ref. 2. A recent study of deep impurities using semiempirical theories has been reported by H.P. Hjalmarson, P. Vogl, D.J. Wolford, and J.D. Dow, Phys. Rev. Lett. **44**, 810 (1980).

5. Hydrogen is not likely to occupy a substitutional site in Si. Here it is used as a representative of sp-bonded $\Delta z=-3$ impurities because of its simplicity and convenience in elucidating the bonding aspects of such impurities.

6. J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. **41**, 895 (1978); Phys. Rev. B **21**, 3545 (1980). [Similar results for Si:V have been obtained by G.A. Baraff and M. Schlüter, Phys. Rev. Lett. **41**, 892 (1978); Phys. Rev. B **19**, 4965 (1979)].

7. Soft-core pseudopotentials were used for both Si (see Ref. 6) and the impurities. For H, the pseudopotential of S. G. Louie [Phys. Rev. Lett. **42**, 476 (1979)] was used. For Zn, a soft-core pseudopotential was constructed treating the d electrons as core electrons. Other soft-core pseudopotentials were also used with no significant changes in the results.

8. For a neutral center, the Si:V level contains two electrons, the Si:H level contains three electrons, and the Si:Zn level contains four electrons. The next impurity in the series is Si:Al ($\Delta z=-1$), which also has a T_2 level in the gap containing five electrons. The latter is described well

by the EMT by treating the T_2 bound state as containing one hole. We note that the EMT point of view for counting electrons remains valid in that Si:Zn has a T_2 level with two bound holes, etc.

9. The wavefunctions also have a canonical shape in k-space. See S.T. Pantelides, N.O. Lipari, and J. Bernholc, Solid St. Comm. **33**, 1045 (1980). Further discussion of the impurity wavefunctions in k-space and comparison with EMT results will be given elsewhere.

10. J.M. Herman III and C.T. Sah, J. Appl. Phys. **44**, 1259 (1973).

11. J.C. Slater, Phys. Rev. **81**, 385 (1951); J.C. Slater, *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York, 1974). We also calculated higher-order corrections to the transition-state formula within the local-density approximation and found them negligible.

12. A Jahn-Teller distortion similar to that of the vacancy [see N.O. Lipari, J. Bernholc, and S. T. Pantelides, Phys. Rev. Lett. **43**, 1354 (1979)] is likely to *increase* the ionization energies because holes to be emitted to the valence band are in the upper, four-fold degenerate state.

13. For the purposes of this work, the Coulombic tail was treated by including basis orbitals on the central atom and two shells of neighbors (see Ref. 6). The uncertainty in the binding energy is estimated to be of order 0.1 eV.

14. J. Bernholc and S.T. Pantelides, Phys. Rev. B **18**, 1780 (1978).

15. For a general critique of tight-binding models see S.T. Pantelides and J. Pollmann, J. Vac. Sci. Technol. **16**, 1349 (1979).

FIGURE CAPTIONS

1. The wavefunction amplitude of the T_2 bound state of Si:V (from Ref. 6), Si:H, and Si:Zn in the (100) plane. The atomic positions are marked by dots. E_B denotes the position of the energy level in the gap.

2. The change in the density of A_1 states for Si:V (solid, from Ref. 6) and Si:H (dashed). Only the solid curve is shown where they can hardly be distinguished.

3. The charge-density change associated with the A_1 resonance of Si:V (from Ref. 6) and Si:H.

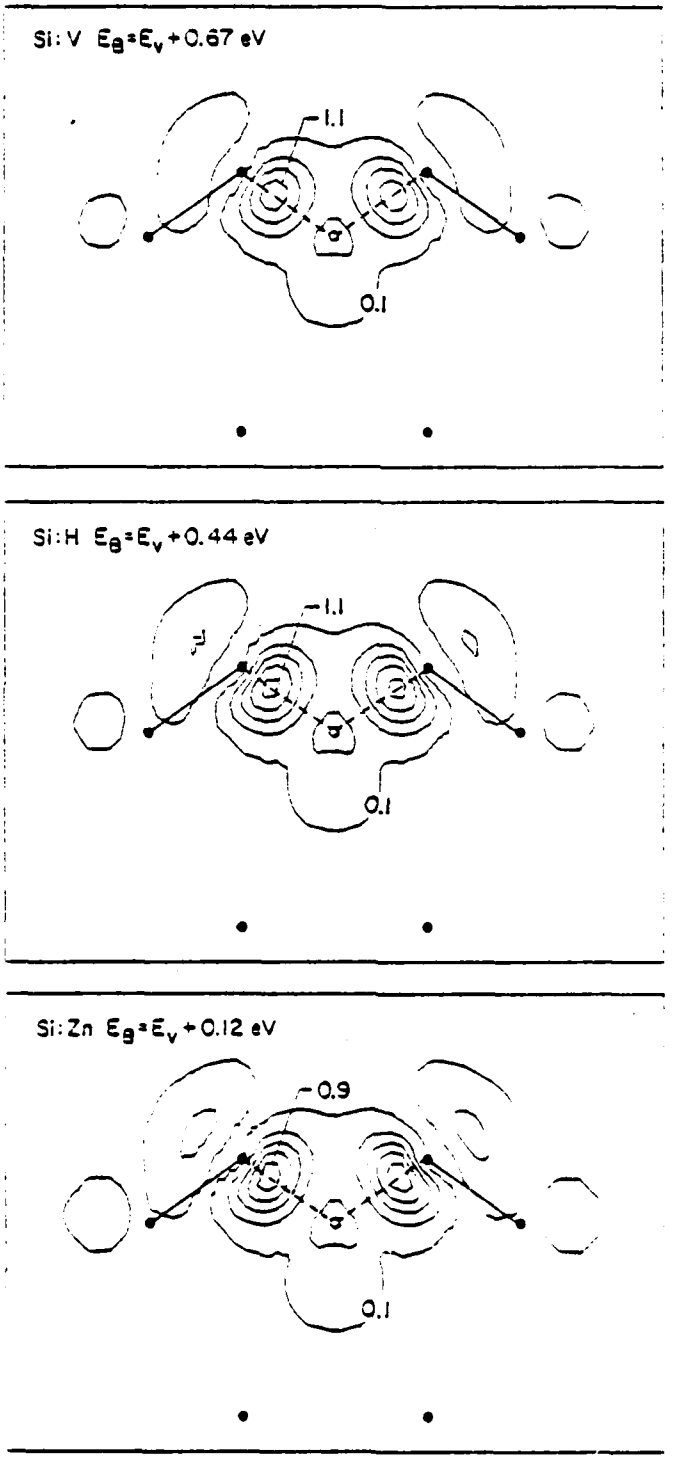

Si:V $E_B = E_v + 0.67$ eV
1.1
σ
0.1
Si:H $E_B = E_v + 0.44$ eV
1.1
π
σ
0.1
Si:Zn $E_B = E_v + 0.12$ eV
0.9
σ
0.1

Fig. 1

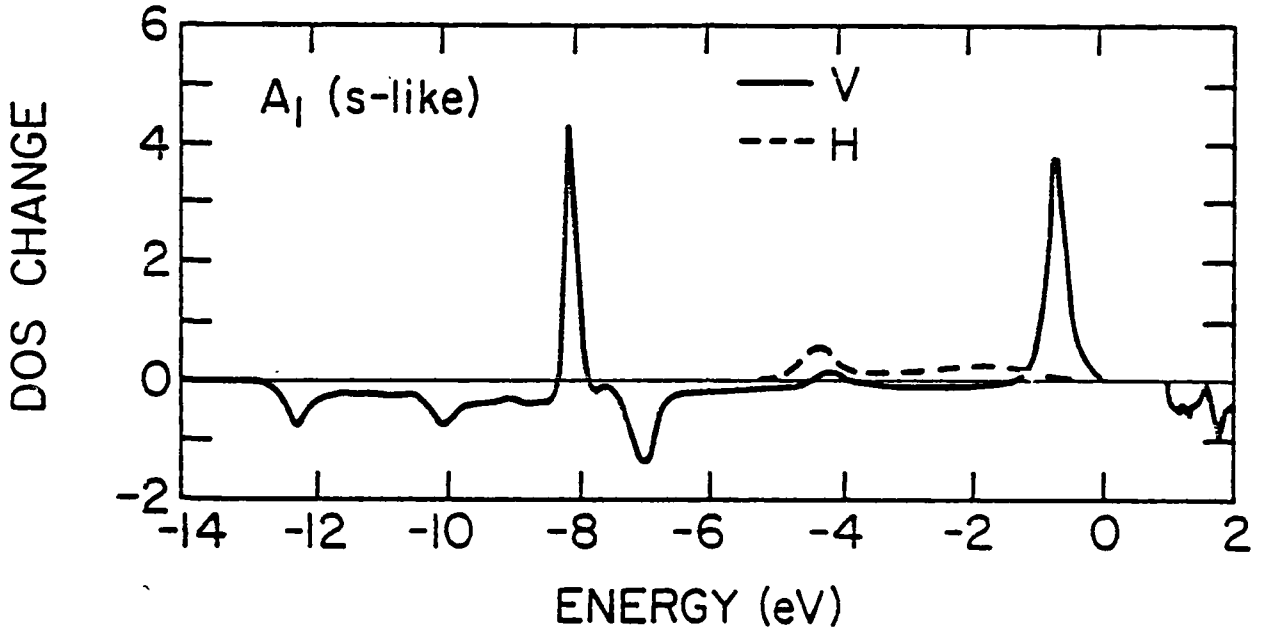


Fig. 2

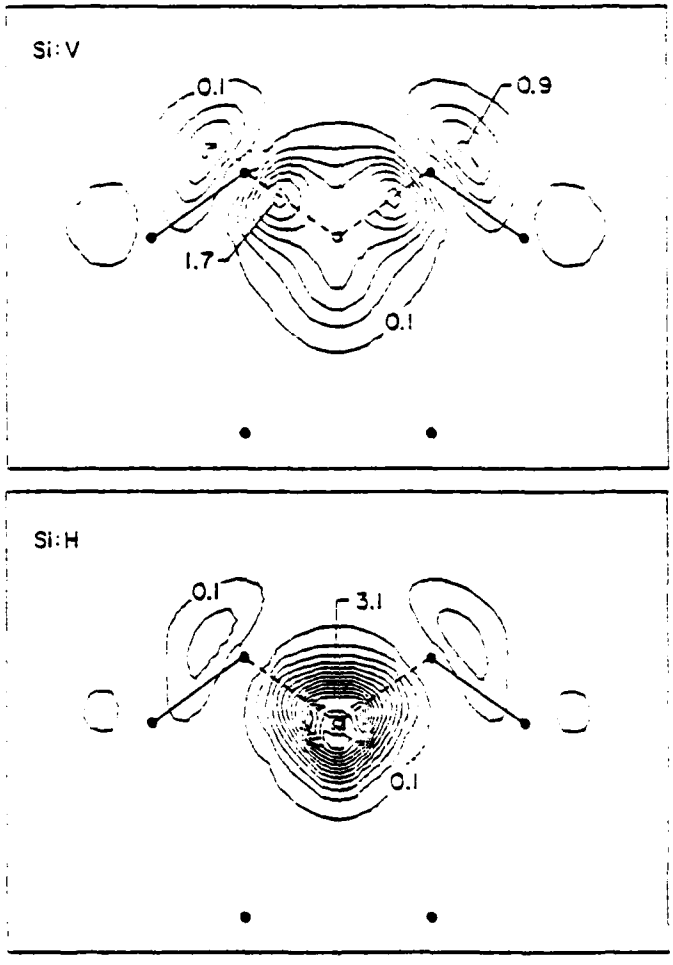


Fig. 3

Appendix N

RC 8252 (#35921) 5/6/80
Solid State Physics 9 pages

FINAL-STATE EFFECTS IN THE EXCITATION SPECTRA OF DEEP IMPURITIES IN SEMICONDUCTORS

Sokrates T. Pantelides

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

and

Hermann G. Grimmeiss

Department of Solid State Physics
Lund Institute of Technology, Lund, Sweden

Typed by Linda P. Rubin (SP.2487)

Abstract: The excitation spectra of deep impurities have usually been interpreted in terms of transitions to continuum states having the same energy distribution and Bloch-like character as the perfect-crystal band states. Here we provide theoretical analysis and experimental evidence showing that deep-level spectra are in fact dominated by bound and quasibound final states induced by the strong short-range impurity potentials.

The excitation spectra of shallow impurities in semiconductors are known to consist of a series of Rydberg-like peaks followed by a continuum[1]. These spectra are well understood in terms of effective-mass theory (EMT).[2] The discrete states below the band edge are the bound excited states of the electron or hole bound to a Coulombic impurity potential. In the EMT treatment, the energies of the band states are assumed to have the same distribution as in the perfect crystal, approximated by an effective-mass parabolic dispersion, whereas the corresponding wavefunctions have been taken to be either perfect-crystal Bloch functions or continuum solutions of the hydrogenic EMT Schrödinger equation[3].

The excitation spectra of deep impurities, on the other hand, are not as well understood. Usually, they exhibit no Rydberg-like discrete states, but consist of a rather structureless curve which rises smoothly and then gradually turns over[4]. Simple models have been developed[5-12] for the continuum part of the spectrum which use different forms of a bound-state wavefunction (e.g., hydrogenic, or, as first proposed by Lucovsky,[5] the bound-state wavefunction of a δ-function impurity potential, or as calculated by a numerical solution of the impurity problem[7,10]). In all these models,[5-12] the final states in the band continuum are approximated as in the case of shallow impurities i.e., the distribution of the final-state energies is taken to be that of the perfect-crystal energy bands (usually effective-mass parabolic dispersions) and the corresponding wavefunctions are taken to be perfect-crystal Bloch functions or, for charged centers, Coulombic continuum solutions. Applications of these models have proved useful in fitting the data and extracting binding energies, but their inability to account for all the data, especially when the observed spectra exhibit any kind of structure, has been noted several times.[3,4,8]

In this paper we focus our attention on the validity of the usual assumption that the final continuum states in deep-level excitation spectra are well approximated by perfect-crystal Bloch functions and energy band dispersions. We will present theoretical and experimental evidence that in many cases this assumption is invalid and that, on the contrary, the dominant

final states are localized or quasilocalized states induced by the strong short-range impurity potential.

We start by addressing the general question of the nature of final states available to a carrier (electron or hole) which, before the excitation takes place, is in a defect-introduced[13] bound state in the gap. We will pursue this analysis in a quasiparticle picture which is capable of taking into account all the essential many-body effects (polarization, electron-hole attraction, etc.) The nature of the available final states depends crucially on the nature of the defect potential. We immediately identify two distinct cases:

(a) If, upon removal of the bound carrier to a propagating continuum state (ionization), the center remains charged, then the defect potential has a screened Coulombic tail and is, therefore, of the form[14]

$$U(r) = U_{SR} + ne^2/\epsilon r, \tag{1}$$

where U_{SR} is a short-range potential, ϵ is the dielectric constant, and n is the net charge at the center upon removal of the bound carrier.

(b) If, upon removal of the carrier, the center is neutral, then the defect potential has no Coulombic tail and consists of only a short-range potential U_{SR}.

We are now in position to distinguish between shallow and deep defects by the strength of U_{SR}. Shallow defects are characterized by a weak U_{SR}. In the absence of a Coulombic tail, as in the case of isovalent impurities, a weak U_{SR} may not even have any bound states. For Coulombic centers, e.g., substitutional group-III and group-V impurities in Si and Ge, a weak U_{SR} represents only a small correction to the Coulombic solutions (chemical shifts). The final states available to a bound carrier are thus Rydberg-like series below band edges and subsidiary minima[15] and continuum states that are slightly modified Bloch functions.

Deep defects, on the other hand, are characterized by strong U_{SR}'s. In some cases, U_{SR} is moderately strong so that U_{SR} and the Coulombic tail are of equal importance. We will not discuss these cases further in this paper. In other cases, however, U_{SR} may be so strong that the Coulombic tail, even if it exists in principle, is only a minor perturbation. One such case is the vacancy in Si, which has recently been investigated by self-consistent calculations.[16-18] These calculations indicate that the vacancy is characterized by a strong, short-range defect potential, which, in addition to producing a deep bound state, causes dramatic changes in the continuum states, e.g., in the form of resonances and antiresonances. It may also introduce additional bound states in the gap[19]. Similar results have since been obtained for deep impurities[20]. Thus, we conclude that, in the case of a deep defect, there exist four types of final states available to a bound carrier:

(i) Rydberg-like series below band edges, when they exist in principle, as discussed above.

(ii) Propagating states which for symmetry or other reasons are unperturbed or slightly-perturbed Bloch states.

(iii) Other, non-Rydberg-like bound states in the gap (such states may or may not exist).

(iv) Strongly modified continuum states such as resonances, which have localized wavefunctions.

Thus, in contrast to shallow impurities whose spectra arise from transitions to Rydberg- and Bloch-like final states, deep-defect spectra may in fact be dominated by transitions to localized states both below and above the band edge, depending on the nature of the process. Yet, such final states have largely been ignored in the analysis of deep-level spectra. For a correct analysis, one must identify the excitation mechanism and corresponding selection rules that will determine the dominant final states. For example, the photoabsorption process would favor transitions to localized states, because optical matrix elements are larger for such states, unless they are forbidden by symmetry[21]. On the other hand, a photoconductivity measure-

ment might fail altogether to detect bound final states unless a secondary process causes carriers in such states to scatter into propagating states. For example, carriers in bound final states may be thermally excited to propagating continuum states (a process known as two-step photothermal ionization[22]). Photoconductivity experiments can also detect quasibound carriers in a resonance because resonances have an intrinsic finite lifetime. The process will depend on the detailed properties of the resonances, the available continuum states (symmetry, wavefunction localization, density of continuum states, strength of coupling to the lattice, etc.), and the availability of other localized states (e.g., hydrogenic shallow states) into which the excited carrier might prefer to decay.

In the remainder of this paper, we will discuss a series of experimental data which illustrate that bound and quasibound final states may indeed play an important role in deep-level spectra.

In Fig. 1 we show the photoionization cross-section for hole emission of the dominant center in Cu-doped ZnSe reported previously by Grimmeiss et al[23]. This spectrum has been measured by a photocapacitance experiment and also by luminescence quenching. In the latter experiment, luminescence is initially obtained from electrons dropping from the conduction bands into the empty trap. This luminescence is then quenched by optically induced hole emission from the trap independent of whether the hole final state is propagating or bound. In Fig. 1(a), the spectrum is plotted on a logarithmic scale over a narrow energy range. The curve looks like a typical deep-impurity spectrum that might lead one to interpret it in terms of the usual simple models that assume transitions to propagating continuum states. In Fig. 1(b) we plot the same spectrum on a linear scale and over a wider energy range, which is possible because ZnSe has a wider gap than Si and most III-V compounds. The presence of a peak[4] is then unmistakable. In order to explore the nature of this peak, photoconductivity measurements have been carried out[25] and the results are shown in Fig. 2. The peak is now absent, indicating that its presence in the photoionization data of Fig. 1 is probably due to transitions

to a bound or quasi-bound state.[26] Even though the photoconductivity spectrum begins at about 0.7 eV and the peak is above 0.8 eV, it is not possible to infer whether the peak is in fact a true bound state or a quasi-resonance within the band continuum because secondary processes can result in an effective photoconductivity threshold below the true band edge.

Such striking experiments are not usually available, but strong evidence can be accumulated that deep-level spectra are often dominated by transitions to bound and quasibound final states induced by the strong short-range potentials. For example, a peak near threshold is also present in the electron-emission photoionization spectrum of the Cu-related center[23] discussed above. The spectrum of Cr-doped GaAs[27] is similar, with a dominant peak near threshold. In that case, virtually identical spectra are obtained by photoabsorption and photoconductivity. It has previously been suggested that the peak is due to a bound state below the band edge and is detected by photoconductivity via a two-step photothermal process (See Ref. 27). Such an interpretation is not consistent with the experimental observation[27] that the peak's height is not strongly reduced at low temperatures. The present analysis suggests that the peak may in fact be a resonance overlapping the continuum which decays into propagating states because of a short intrinsic lifetime.

Another example that may exhibit the effects discussed in this paper is GaP:O. Recently, Samuelson and Monemar[11] showed that the observed hole-emission cross section of GaP:O, which appears to be a broad featureless hump, actually exhibits two distinct peaks when phonon broadening is removed by a deconvolution procedure. The temperature dependence of the two peaks was found to be markedly different. It was experimentally determined that the two peaks are due to the same center and the authors speculated that isolated O in GaP may have two deep levels in the gap separated by about 80 meV. The present analysis opens up the possibility that the temperature-dependent peak is a final-state resonance, but further work is necessary to elucidate the structure of this center.

Finally, we mention the case of a phosphorus impurity occupying an oxygen site in SiO_2. This system has been previously analysed[28] and the peak just above the photoionization threshold was identified to be a quasi-bound resonance similar in character to the excitonic resonance that dominates the optical spectrum of the pure material.

In summary, we have presented theoretical and experimental evidence that deep-level excitation spectra are likely to be dominated by transitions to bound and quasibound states both below and above the continuum threshold. These results call for a reexamination of deep-level spectra that have thus far been interpreted in terms of transitions to propagating continuum states.

Acknowledgement: Part of this work was done while one of us (S.T.P.) was on sabbatical leave at the University of Lund, Sweden, in 1978. Some of the experiments were done while HGG was visiting the Philips Laboratory at Briarcliff Manor, New York. The work was supported in part by the Air Force Office of Scientific Research under Contract No. F49620-77-C-0005. We acknowledge valuable discussions with U. von Barth, J. Bernholc, N. Kullendorf, and T. N. Morgan.

REFERENCES

1. See e.g., R.L. Aggarwal and A.K. Ramdas, Phys. Rev. 140 A246 (1965); J.H. Reuszer and P. Fisher, Phys. Rev. 135 A1125 (1964) Rydberg series are sometimes also observed below the extrema of higher bands [e.g. S. Zwerdling, B. Lax, K.J. Button, and L.M. Roth, Phys. Rev. Lett. 4, 173 (1960)] and below subsidiary extrema of the nearest band.
2. W. Kohn, Solid State Phys. 5, 257 (1957); For more recent reviews see S.T. Pantelides, Rev. Mod. Phys. 50, 797 (1978), and M. Altarelli and F. Bassani, Handbook of Semiconductor Physics, ed. by S. Keller, to be published.
3. See Pantelides, Ref. 2 for original references and the different merits of the two choices.
4. For recent reviews, see H.G. Grimmeiss, Ann. Rev. Material. Sci. 7, 341 (1977); G.L. Miller, D.V. Lang, and L.C. Kimerling, Ann. Rev. Material. Sci., 7, 377 (1977).
5. G. Lucovsky, Solid State Comm. 3, 299 (1965).
6. H.B. Bebb and R.A. Chapman, J. Phys. Chem. Solids 28, 2087 (1967).
7. T.H. Ning and C.T. Sah, Phys. Rev. B. 4 3482 (1971).
8. H.G. Grimmeiss and L.Å. Ledebo, J. Phys. C. 8, 2615 (1975).
9. S.T. Pantelides and J. Bernholc, Proceedings of the 9th International Conference on Defects and Radiation Effects in Semiconductors, Dubrovnik, Yugoslavia 1976, (Inst. of Phys. London, 1977), p. 465.
10. M. Jaros, J. Phys. C. 8, 2455 (1975).
11. B. Monemar and L. Samuelson, Phys. Rev. B 18, 809 (1978); L. Samuelson and B. Monemar, Phys. Rev. B 18, 830 (1978).
12. B.K. Ridley, Solid State Electron. 21, 1319 (1978).
13. We generalize our discussion by using the term defect instead of impurity since it applies to more general types of localized defects (vacancies, vacancy-impurity complexes, etc.)

14. Defect potentials have multipole-field tails of the form $r^{-(\ell+1)}$ with $\ell > 0$ whose effect is negligible [see, e.g., L.J. Sham, Phys. Rev. 150, 720 (1966).

15. Rydberg-like states below band edges always exist in principle in the presence of a Coulombic potential. [For a proof, see N.F. Mott and R.W. Gurney, Electronic Processes in Ionic-Solids, (Clarendon, Oxford, 1940)].

16. J. Bernholc, N.O. Lipari, and S.T. Pantelides, Phys. Rev. Lett. 40, 892 (1978).

17. G.A. Baraff and M. Schlüter, Phys. Rev. Lett. 40, 895 (1978).

18. In the calculations of Refs. 16 and 17, the vacancy potential does not have a Coulombic tail because of the local-density approximation to the density-functional formalism of the one-electron problem. The vacancy potential does, in principle, have a Coulombic tail, but its inclusion in the calculation would have a small effect on the final results [J. Bernholc, N.O. Lipari, and S.T. Pantelides, unpublished].

19. E.g., the tetragonally distorted vacancy has a doubly-degenerate bound state in the lower part of the gap containing two electrons and an empty quadruply-degenerate bound state in the upper gap [N.O. Lipari, J. Bernholc, and S.T. Pantelides, Phys. Rev. Lett. 43, 1354 (1979).

20. J. Bernholc, S.T. Pantelides, N.O. Lipari, and A. Baldereschi, to be published.

21. E.g., in the case of the vacancy (Refs. 16, 17) one finds a p-like bound state in the gap and an s-like resonance just below the valence-band top. Clearly, optical matrix elements for hole emission would be large for bound-to-resonance transitions and small for bound-to-band transitions (the Bloch functions at the top of the valence bands are p-like).

22. J.M. Lifshitz and M.I. Kaganov, Usp. Fiz. Nauk 69, 419 (1959) [Sov. Phys. - Usp. 2, 831 (1960)]; J.M. Lifshitz and F. Ya Nad, Sov. Phys. - Dokl. 10, 532 (1965).

23. G.H. Grimmeiss, C. Ovrén, W. Ludwig and R. Mach, J. Appl. Phys. 48, 5122 (1977).

24. This peak has been discussed by C. Ovrén [Doctoral thesis, University of Lund, unpublished], who suggested that it may be due to a discrete state below the band

edge whose origins were not well understood. The remainder of the spectrum was attributed by Ovrén to transitions to Bloch states in the band continuum.

25. K. Kosai, H.G. Grimmeiss, B.J. Fitzpatrick, and R.N. Bhargava, Philips Laboratories Tech. Rep. No. 308 (1978), unpublished.

26. The resonance may be due to Cu d states. Recently, R.N. Bhargava has suggested that the 0.7 eV center is unlikely to be simple substitutional Cu. If it is an unknown Cu-related complex, it is not possible at this stage to deduce detailed information about the nature of the states involved. It is likely, however, to involve pairing with a shallow acceptor, which would provide localized states into which a quasi-bound hole might decay.

27. D. Bois and P. Pinard, Japan J. Appl. Phys. 12, 936 (1973).

28. R.F. DeKeersmaecker, D.J. DiMaria, and S.T. Pantelides, in The Physics of SiO_2 and Its Interfaces, ed. by S.T. Pantelides, (Pergamon, New York, 1978), p. 189.

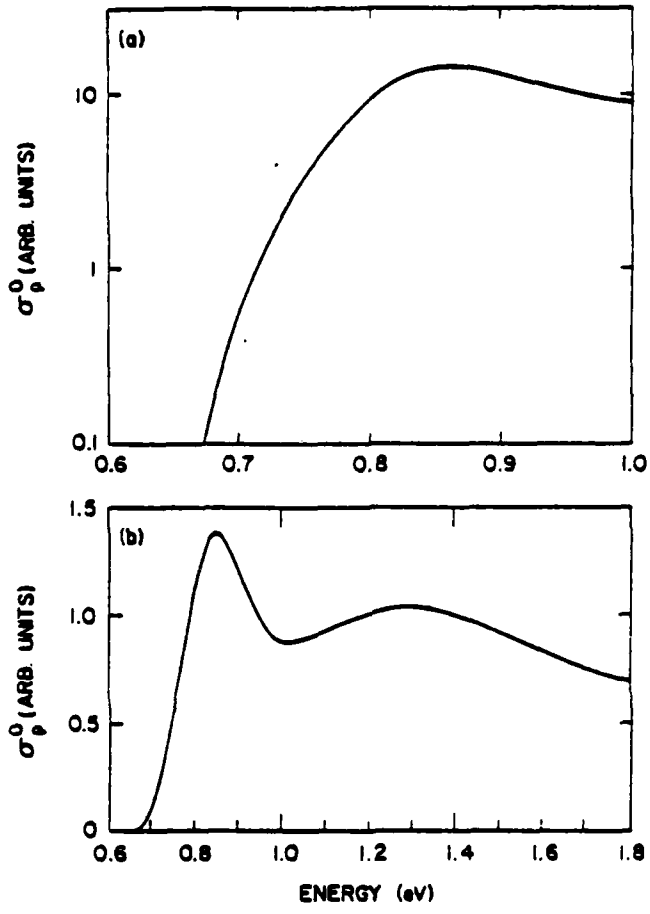


Fig. 1. The photoionization spectrum of ZnSe:Cu as measured by photocapacitance (this work, Ref. 25) and also by luminescence quenching (Ref. 23). In (a), a logarithmic scale is used in the ordinate, whereas in (b) a linear scale is used. The two plots are over different energy ranges. See text.

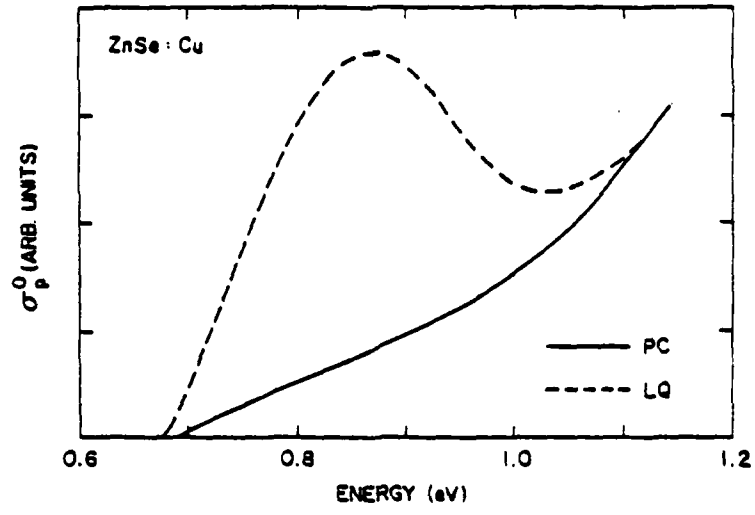


Fig. 2. The photoionization spectrum of ZnSe:Cu as measured by luminescence quenching (dashed line, Ref. 23) and by photoconductivity (solid line, this work, Ref. 25). The ordinate is a linear scale.